

(1) International Patent Classification 6 : C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, C12N 1/21, C07K 14/47, 16/18, C12Q 1/68, G01N 33/50, 33/53, 33/68, A61K 38/17	(11) International Publication Number: <b>WO 98/39448</b>
(43) International Publication Date: 11 September 1998 (11.09.98)	<b>A2</b>

(21) International Application Number:	PCT/US98/04693
(22) International Filing Date:	6 March 1998 (06.03.98)
(30) Priority Data:	
600/40,162	7 March 1997 (07.03.97) US
600/40,333	7 March 1997 (07.03.97) US
600/38,621	7 March 1997 (07.03.97) US
600/40,161	7 March 1997 (07.03.97) US
600/40,656	7 March 1997 (07.03.97) US
600/40,334	7 March 1997 (07.03.97) US
600/40,336	7 March 1997 (07.03.97) US
600/40,163	7 March 1997 (07.03.97) US
600/43,580	11 April 1997 (11.04.97) US
600/43,568	11 April 1997 (11.04.97) US

(Continued on the following page)

MD 20874 (US), DUAN, Roxanne (US/US); 4541 Fairfield Drive, Beltsville, MD 20814 (US), HU, Jing-Shan (CN/US); 1247 Lakeside Drive #3034, Sunnyvale, CA 94066 (US) US/US); MOORE, Paul, A. (GB/US); Apartment #104, 200831 Holy Ridge Drive, McLean, VA 22102 (US), SHI, Yongfang (US), 437 West Side Drive, Gaithersburg, MD 20878 (US), LARLEYBUR, David, W. (US/US); 1615 Q Street, N.W. #807, Washington, DC 20009 (US), LI, Yi (CN/US); 12474 Lakeside Drive #3034, Sunnyvale, CA 94066 (US), ZENG, Zhishen (CN/US); 13950 Saddleview Drive, Gaithersburg, MD 20878 (US), KYAW, Hla (BU/US); 520 Sengnashin Circle, Frederick, MD 21703 (US), US).

***(Continued on the following page)***

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]: 9410 Key West Avenue, Rockville, MD 20850 (US).

(74) Agent: BROOKES, Anderson, A. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US).

**(72) Inventors; and**

(75) *Inventors/Applicants for US only:* RUBEN, Steven, M. (US/US); 18328 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig A. (US/US); 22400 Rolling Hills Road, Laytonville, MD 20862 (US). FISCHER, Carole L. (US/US); 5810 Hall Street, Burke, VA 22015 (US). SORPET, Daniel R. (US/US); 15050 Stillfield, Place, Centerville VA 22020 (US). CARTER, Kenneth C. (US/US); 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). BEDNARIK, Daniel P. (US/US); 8822 Blue Sea Drive, Columbia, MD 21046 (US). ENDRES, Gregory A. (US/US); 9729 Caguel Fern Drive, Potomac, MD 20854 (US). YU, Guo-Liang (CN/US); 13324 Starw Blue Lane, Darnestown, MD 20878 (US). NI, Jian (CN/US); 5502 Manorfield Road, Rockville, MD 20853 (US). FENG, Ping (CN/US); 4 Relda Court, Gaithersburg, MD 20878 (US). YOUNG, Paul E. (US/US); 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John M. (US/US); 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRER, Ann, M. (US/US); 13203 L Astoria Hill Court, Germantown,

Published

**Without international search report and to be republished upon receipt of that report.**

**(54) Title: 186 HUMAN SECRETED PROTEINS**

### (57) Abstract

The present invention relates to 185 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

**FOR THE PURPOSES OF INFORMATION ONLY**

**Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT**

60004.319	11 April 1997 (11.04.97)	US
60004.314	11 April 1997 (11.04.97)	US
60004.311	11 April 1997 (11.04.97)	US
60004.671	11 April 1997 (11.04.97)	US
60004.674	11 April 1997 (11.04.97)	US
60004.669	11 April 1997 (11.04.97)	US
60004.312	11 April 1997 (11.04.97)	US
60004.313	11 April 1997 (11.04.97)	US
60004.672	11 April 1997 (11.04.97)	US
60004.578	11 April 1997 (11.04.97)	US
60004.576	11 April 1997 (11.04.97)	US
60004.670	11 April 1997 (11.04.97)	US
60004.600	23 May 1997 (23.05.97)	US
60004.615	23 May 1997 (23.05.97)	US
60004.597	23 May 1997 (23.05.97)	US
60004.592	23 May 1997 (23.05.97)	US
60004.633	23 May 1997 (23.05.97)	US
60004.583	23 May 1997 (23.05.97)	US
60004.617	23 May 1997 (23.05.97)	US
60004.618	23 May 1997 (23.05.97)	US
60004.504	23 May 1997 (23.05.97)	US
60004.592	23 May 1997 (23.05.97)	US
60004.584	23 May 1997 (23.05.97)	US
60004.581	23 May 1997 (23.05.97)	US
60004.500	23 May 1997 (23.05.97)	US
60004.587	23 May 1997 (23.05.97)	US
60004.594	23 May 1997 (23.05.97)	US
60004.598	23 May 1997 (23.05.97)	US
60004.613	23 May 1997 (23.05.97)	US
60004.582	23 May 1997 (23.05.97)	US
60004.586	23 May 1997 (23.05.97)	US
60004.612	23 May 1997 (23.05.97)	US
60004.632	23 May 1997 (23.05.97)	US
60004.601	23 May 1997 (23.05.97)	US
60004.595	23 May 1997 (23.05.97)	US
60004.599	23 May 1997 (23.05.97)	US
60004.588	23 May 1997 (23.05.97)	US
60004.585	23 May 1997 (23.05.97)	US
60004.586	23 May 1997 (23.05.97)	US
60004.590	23 May 1997 (23.05.97)	US
60004.589	23 May 1997 (23.05.97)	US
60004.593	23 May 1997 (23.05.97)	US
60004.614	23 May 1997 (23.05.97)	US
60004.501	23 May 1997 (23.05.97)	US
60004.504	06 June 1997 (06.06.97)	US
60004.964	06 June 1997 (06.06.97)	US
60004.610	06 June 1997 (13.06.97)	US
60005.1926	08 July 1997 (08.07.97)	US
60005.874	16 July 1997 (16.07.97)	US
60005.72	18 August 1997 (18.08.97)	US
60005.886	22 August 1997 (22.08.97)	US
60005.877	22 August 1997 (22.08.97)	US
60005.889	22 August 1997 (22.08.97)	US
60005.893	22 August 1997 (22.08.97)	US
60005.630	22 August 1997 (22.08.97)	US
60005.878	22 August 1997 (22.08.97)	US
60005.662	22 August 1997 (22.08.97)	US
60005.872	22 August 1997 (22.08.97)	US
60005.882	22 August 1997 (22.08.97)	US
60005.637	22 August 1997 (22.08.97)	US
60005.903	22 August 1997 (22.08.97)	US
60005.888	22 August 1997 (22.08.97)	US
60005.879	22 August 1997 (22.08.97)	US
60005.880	22 August 1997 (22.08.97)	US
60005.894	22 August 1997 (22.08.97)	US
60005.911	22 August 1997 (22.08.97)	US
60005.635	22 August 1997 (22.08.97)	US
60005.914	22 August 1997 (22.08.97)	US
60005.894	22 August 1997 (22.08.97)	US
60005.864	22 August 1997 (22.08.97)	US
60005.631	22 August 1997 (22.08.97)	US
60005.845	22 August 1997 (22.08.97)	US
60005.892	22 August 1997 (22.08.97)	US
60005.632	22 August 1997 (22.08.97)	US
60005.664	22 August 1997 (22.08.97)	US
60005.871	22 August 1997 (22.08.97)	US
60005.881	22 August 1997 (22.08.97)	US
60005.909	22 August 1997 (22.08.97)	US
60005.872	22 August 1997 (22.08.97)	US
60005.862	22 August 1997 (22.08.97)	US
60005.887	22 August 1997 (22.08.97)	US
60005.908	22 August 1997 (22.08.97)	US
60005.907	22 August 1997 (22.08.97)	US
60005.761	05 September 1997 (05.09.97)	US
60005.650	05 September 1997 (05.09.97)	US
60005.785	12 September 1997 (12.09.97)	US
60006.106	02 October 1997 (02.10.97)	US

AL	Albania	BS	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AN	Andorra	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SE	Sweden
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	SD	Sudan
BD	Bangladesh	GE	Georgia	MD	Republic of Moldova	SG	Singapore
BE	Belgium	GR	Greece	MG	Madagascar	SI	Sri Lanka
BF	Burkina Faso	HN	Honduras	ML	The former Yugoslav Republic of Macedonia	TH	Thailand
BG	Bulgaria	HU	Hungary	MT	Malta	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MY	Malaysia	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VE	Venezuela
CG	Congo	KE	Kenya	NL	Netherlands	VN	Viet Nam
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	YU	Yugoslavia
CI	Cote d'Ivoire	KO	Democratic People's Republic of Korea	NZ	New Zealand	ZW	Zimbabwe
CM	Cameroon	RP	Philippines	PT	Portugal		
CN	China	KR	Republic of Korea				
CU	Cuba	KZ	Kazakhstan				
CZ	Czech Republic	LC	Saint Lucia	RO	Romania		
DE	Germany	LI	Liechtenstein	RU	Russian Federation		
DK	Denmark	LK	Sri Lanka	SD	Sudan		
EE	Estonia	LT	Lithuania	SE	Sweden		
		LU	Luxembourg	SG	Singapore		

## 186 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

### *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

### *Detailed Description*

#### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 12301 Park Lawn Drive, Rockville, Maryland 20852, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA contained within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA, followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing) or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polynucleotide or polynucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, uridyated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods.

Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotyrosine/inositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, **PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES**, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); **POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS**, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., *Meth Enzymol* 182:626-646 (1990); Rattan et al., *Ann NY Acad Sci* 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table I.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## 25 Polynucleotides and Polypeptides of the Invention

### **FEATURES OF PROTEIN ENCODED BY GENE NO: 1**

This gene is expressed primarily in testes tumor and to a lesser extent in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly of the testes, and defects of the central nervous system such as seizure and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly cancer of the testes and central nervous system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, brain and other tissue of the nervous system, and blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of testicular cancer and treatment of central nervous system disorders since this gene is primarily expressed in the testes tumor and developing brain.

### **FEATURES OF PROTEIN ENCODED BY GENE NO: 2**

This gene is expressed primarily in cancer tissues, such as breast cancer and Wilm's tumor, and to a lesser extent in fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and/or tumors, particularly, those found in the breast, and developmental abnormalities or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and fetal tissue and, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 314 as residues: Pro-11 to Thr-18, Leu-43 to Pro-50, Gly-64 to Leu-72, and Leu-81 to Lys-86.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of cancers and/or tumors, particularly, those found in the breast since expression is mainly in cancer/tumor tissues. May serve as therapeutic proteins for proliferation/differentiation of fetal



**FEATURES OF PROTEIN ENCODED BY GENE NO: 3**

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in spleen, chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or leukemias, diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment/diagnosis of blood disorders or leukemias, diseases of the immune system since expression is in tissues related to immune function.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 4**

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or lymphocytic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 5**

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 317 as residues: Pro-13 to Lys-21.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 6**

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 318 as residues: Lys-31 to Lys-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood diseases since it is expressed in tissues related to immune function.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 7

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system and brain associated diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders, immune diseases or brain associated diseases (specifically of the pineal gland) since expression is in tissues related to immune function.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 8

The translation product of this gene shares sequence homology with an organic cation transporter which is thought to be important in organic cation uptake in the kidney and liver. (See Accession No. 2343059.) Preferred polypeptide fragments comprise the amino acid sequence ITTAIQMCLVNXELYPTFVRNXGVMVCSLDCIGGIPT FIVRLRVVQALPLFVLGALLAAGVTLLPETKGVLPETMKDAENLGRKAKPKENTYLLK VQTSEPSGT (SEQ ID NO: 615) or TMKDAENLGRKAKPKENT (SEQ ID NO: 616) as well as N-terminal and C-terminal deletions of these fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic and renal diseases where drug elimination/cation exchange (organic cation uptake) in the liver and kidney are problematic. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 320 as residues: Asn-64 to Asn-74, and Gln-81 to Gly-87.

The tissue distribution and homology to organic cation transporter indicate that polynucleotides and polypeptides corresponding to this gene are useful as a polyspecific transporter that is important for drug elimination in the liver (and possibly kidney) since expression is found in the liver.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 9

This gene is expressed primarily in eosinophil induced with IL-5 and to a lesser extent in fetal liver and spleen. This gene also maps to chromosome 15, and therefore can be used in linkage analysis as a marker for chromosome 15.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system, particularly allergies or asthma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the

standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosis of diseases involving eosinophil reactions since expression seems to be concentrated in eosinophils and other tissues involved in immunity such as the liver and spleen.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 10

This gene is expressed primarily in tissues of hematopoietic lineage and to a lesser extent in Hodgkins lymphoma. Any frame shifts in this sequence can easily be clarified using known molecular biology techniques.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and immune deficiency or dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, lymphoid and reticuloendothelial tissues, and cancerous tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/ diagnosis for lymphomas or immune dysfunction or as a therapeutic protein useful in immune modulation based on expression in anergic T-cells and lymphomas.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in neutrophils and to a lesser extent in activated lymphoid cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the cell type present in a biological sample and for diagnosis of diseases and conditions: inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 323 as residues: Glu-40 to Lys-46.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for modulation of an immune reaction or as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in brain and to a lesser extent in activated T-cells. It is likely that the open reading frame containing the predicted signal peptide continues in the 5' direction. Preferred polypeptide fragments comprise the amino acid sequence PRVRNSPEDGLSLTGDSCKL (SEQ ID NO:617).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disorders including ischemic shock, alzheimers and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain, and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 324 as residues: Ser-5 to Glu-14, Ile-21 to Pro-35, Ser-65 to Asp-81, Cys-89 to Val-96, Lys-136 to Ser-145, Ile-152 to Met-169, and Arg-189 to Lys-196.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic/treatment for cancers of the given tissue or in the treatment of neurological disorders of the CNS.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 13**

This gene was also recently cloned by other groups, naming this calcium-activated potassium channel gene, hKC4. (See Accession No. AF033021, see also, Accession No. 2584866.) This gene is mapped to human chromosome 19q13.2. A second signal sequence likely exists upstream from the predicted signal sequence as described in

Table 1. Preferred polypeptide fragments comprise: QADDLQATVAALCVLRGGGPWAG  
 SWLSPKTPGAMGGDLVLGLGALRRRKRL (SEQ NO: 618); or EQEKSAGWALVLAXXGIGL  
 MVLHAEMLVFGGCSAVNATGHLSDTLVLPIPTFLTIGYGDVVPCTMWGKIVCLCTGVMGVCC  
 TALLVAVVARKLEFNKAEKHVHFMMDIQYTKEMKESAAARVLQEAWMFYKHTRRKESHAAR  
 XHQRXLLAAINAFRQVRLKHKRLREQVNSMVDISKMHMILYDLQQLSSSHRALEKQIDTLTG  
 KLDALTELLSTALGPRQLPEPSQSK (SEQ ID NO: 619), as well as N-terminal and C-  
 terminal deletions. Also preferred are polynucleotide fragments encoding these  
 polypeptide fragments.

This gene is expressed primarily in breast lymph node and T-cells, and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematologic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue, blood cells and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 325 as residues: Arg-13 to Lys-23.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment/diagnosis of hematologic and diseases involving immune modulation based on distribution in the lymph node and T-cells.

35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 14**

This gene was recently cloned by another group, calling it PAPS synthase.

(See Accession No. e1204135.) Preferred polypeptide fragments comprise the amino acid sequence YQAHVSRNRKRGQVVGTGRGFRGCTVWLTOLSGAGK (SEQ ID NO: 620).

Also preferred are the polynucleotide fragments encoding this polypeptide fragment. It has been discovered that this gene is expressed primarily in benign prostate hyperplasia, Human Umbilical Vein Endothelial Cells and to a lesser extent in smooth muscle and Human endometrial stromal cells-treated with estradiol.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammation, ischemia, and restenosis, based on endothelial cell and smooth muscle cell expression, and prostate diseases such as benign prostate hyperplasia or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate or vessels of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, endothelial cells, smooth muscle, and endometrium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 326 as residues: Arg-21 to Asp-26, Lys-35 to Lys-44, Glu-49 to Asn-58.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosing diseases or conditions where the endothelial cell lining of the veins and arteries of underlying smooth muscle are involved.

30

**FEATURES OF PROTEIN ENCODED BY GENE NO: 15**

This gene is expressed primarily in human 6 week embryo and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental anomalies or fetal deficiencies. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly developmental in nature, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 327 as residues Lys-50 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of developmental abnormalities.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 16

15 This gene is expressed primarily in kidney and amygdala and to a lesser extent in fetal tissues. This gene is mapped to chromosome 14, and therefore is useful in linkage analysis as a marker for chromosome 14.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) present in a biological sample and for diagnosis of diseases and conditions: kidney diseases, neurological disorders and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s). For a number of disorders of the above tissues, particularly of the renal system or developing fetal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, amygdala, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of conditions affecting the brain, kidneys and fetal development.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 17

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: solid tumors similar to ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 329 as residues Ser-51 to Val-56.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of solid tumors of the reproductive system such as ovarian cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed primarily in brain medulloblastoma. Preferred polypeptide fragments comprise the amino acid sequence: IRHEQHPNFSLEMSKSSLLFLPQLILLPVCAHLHELNC (SEQ ID NO: 643) and SPTSESKGHLLDQAEHRHVYAGALVGVSGGLTLTTCGSGPTKPRATKNYFLKRLLDQEMHRIAN (SEQ ID NO: 644), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the CNS or. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating medulloblastoma or similar tumors.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 19

This gene is expressed primarily in adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose tissues expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating obesity by regulating the function and number of adipocytes

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in B cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of the immune system with an emphasis on B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumors of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

35

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of B cell derived tumors based on its expression in b cell lymphomas

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene is expressed primarily in immune cells and to a lesser extent in fetal tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cells of the immune system; and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:333 as residues Asp-10 to Pro-19, Ser-74 to Tyr-79, Glu-95 to Lys-110.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of diseases involving alterations in T cell activity.

25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 22

It has been discovered that this gene is expressed primarily in ovarian tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the ovary. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumors of the reproductive organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian

30

35

and other reproductive tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 334 as residues: Leu-22 to Gln-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovarian tumors as it has only been identified in ovarian tumors.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 23

It has been discovered that this gene is expressed primarily in fetal tissues and to a lesser extent in osteoclastoma cell line

Therefore, polynucleotides and polypeptides of the invention are useful as

15

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: osteoporosis or arthritis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment of conditions of abnormal bone remodeling due to enhanced activity of osteoclasts. This may be useful as a specific marker for malignancies derived from osteoclasts or their precursors.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 24

The translation product of this gene shares sequence homology with a periplasmic ribonuclease which is thought to be important in degrading extracellular polynucleotides

35

It has been discovered that this gene is expressed primarily in serum treated smooth muscle cells

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular disease such as restenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculature expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Gln-30 to Lys-36, and Pro-41 to Arg-48.

15

The tissue distribution and homology to ribonucleases indicate that

polynucleotides and polypeptides corresponding to this gene are useful for treatment of pathological conditions of smooth muscle associated with bacterial or viral infiltration

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 25

20

This gene is expressed primarily in Early Stage Human Brain.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain development and related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

25

identification of the tissue(s) or cell type(s). For a number of disorders of the above

tissues or cells, particularly of the human brain development and related diseases,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

30

fluid or spinal fluid) or another tissue or cell sample taken from an individual having

such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to this gene indicate that polynucleotides

35

and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases affecting human brain development and related diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 26**

It has been discovered that this gene is expressed primarily in human brain tissue.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases and other diseases related to brain diseases, which may be caused by brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

10 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

15 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain diseases and other diseases related.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 27**

It has been discovered that this gene is expressed primarily in Anergic T-cells.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases, inflammatory diseases and diseases related to T lymph cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

30 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune diseases, inflammatory diseases and diseases related to T lymph cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell

types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

35

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for immune diseases, inflammatory diseases and diseases related to T lymph cells.

5

**FEATURES OF PROTEIN ENCODED BY GENE NO: 28**

The translation product of this gene shares sequence homology with *Shigella flexneri* positive transcriptional regulator CriR (criR) gene which is thought to be important in regulation of gene expression.

10

This gene is expressed primarily in human synovial sarcoma and normal human brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases particularly sarcomas of the synovium. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above

15 tissues or cells, particularly of the human brain and synovium and other related human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., synovial tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human synovial sarcoma and other related human brain diseases.

25

**FEATURES OF PROTEIN ENCODED BY GENE NO: 29**

This gene is expressed in bone marrow, infant brain, fetal liver and spleen, prostate and to a lesser extent in pineal gland, adipose tissue, kidney, adrenal gland, umbilical vein endothelial cells, and T cells.

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases related to bone marrow or

35



hematopoietic tissues, prostate, kidney, adrenal gland, and cardiovascular tissue or organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases related to hematopoietic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, hematopoietic cells, pineal gland, adipose tissue, kidney, adrenal gland, endothelial cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases related to hematopoietic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 30

This gene is expressed primarily in meningea and to a lesser extent in breast and adult brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases of the meningea and related brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the meningea and related brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., meningea, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the meningea and related brain diseases.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 31

This gene is expressed in meningea, fetal spleen, osteoblast and to a lesser extent in activated T-cells, endometrial stromal cells, fetal lung, HL-60, thymus, testis and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: meningeal disease, osteoporosis, immune diseases, and hematopoietic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the meningeal diseases, osteoporosis, immune diseases, and hematopoietic diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, endometrium, lung, thymus, testis, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of meningeal, osteoporosis, immune diseases, hematopoietic diseases, testis diseases and lung diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 32

This gene is expressed primarily in human thymus and to a much lesser extent in infant brain, T-cells, smooth muscle, endothelial cells, bone marrow, human ovarian tumor and keratinocytes, osteoclastoma, breast, and tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases involving the thymus, particularly thymic cancer and diseases involving T-cell maturation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells, particularly of the thymus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, brain, and other tissue of the nervous system, blood cells, bone marrow, ovaries, and testes, and other reproductive tissue, mammary tissue, tonsils, melanocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the thymus particularly thymic cancer and diseases involving T-cell maturation.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is expressed primarily in human tonsils, and placenta, and to a lesser extent in adipocytes, melanocyte, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases, immune diseases, and obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inflammatory diseases, immune diseases, and obesity, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., tonsils, placenta, adipocytes, melanocytes, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to this gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases such as inflammation, immune diseases, and obesity.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene is expressed in activated T cells, and to a lesser extent in pituitary, testis, and breast lymph node.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases relating to T cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary, testes and other reproductive tissue, mammary tissue, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of immune disorders.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 35

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain, and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 36**

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in human ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the ovarian disorders such as those involving germ cells, ovarian follicles, stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovariohpaty.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 38**

This gene is expressed primarily in lymph node breast cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for used as a diagnostic marker for breast cancer.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 39**

This gene is expressed primarily in brain and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal disorders such as trauma, brain degeneration, and brain tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of neuronal disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 40

This gene is expressed in early stage human embryo, adrenal gland tumor, and immune tissues such as fetal liver, fetal spleen, T-cell, and myeloid progenitor cell line and to a lesser extent in ovary, colon cancer, and a few other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis including adrenal gland tumor, colon cancer and various other tumors, developmental and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, early stage human tissues, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, blood cells, bone marrow, ovary and other reproductive tissue, and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of immune and developmental disorders, and tumorigenesis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene is expressed primarily in fetal lung, endothelial cells, liver, thymus and a few other immune tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders such as immune deficiency and autoimmune diseases, pulmonary diseases, liver diseases, and tumor metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal lung, liver, endothelial cells, and immune tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain

tissues and cell types (e.g., lung, endothelial cells, liver, thymus, and other tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of immune disorders and pulmonary and hepatic diseases. Its promoter may also be used for immune system and lung-specific gene therapies. The expression of this gene in endothelial cells indicates that it may also involve in angiogenesis which therefore may play role in tumor metastasis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 42

This gene is expressed primarily in liver, thyroid, parathyroid and to a lesser extent in fetal lung, stomach and early embryos.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic regulation, obesity, hepatic failure, hepatocellular tumors or thyroiditis and thyroid tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive/endocrine system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, thyroid, parathyroid, lung, stomach, and embryonic tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and the extracellular locations indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of digestive/endocrine disorders, including metabolic regulation, hepatic failure, malabsorption, gastritis and neoplasms.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 43**

This gene is expressed primarily in Schizophrenic adult brain, pituitary, front cortex, hypothalamus and to a lesser extent in retina, adipose and stomach cancer and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal tissue, adipose, stomach, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nerve system, including schizophrenia, neurodegeneration, and neoplasia. Additionally, a secreted protein in brain may serve as an endocrine.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 44**

The translation product of this gene shares sequence homology with GTP binding proteins which are thought to be important in signal transduction and protein transport.

This gene is expressed primarily in umbilical vein and microvascular endothelial cells, GM-CSF treated macrophage, anergic T cells, osteoblast, osteoclast, CD34+ cells and to a lesser extent in gall bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: bone formation and growth, osteonecrosis, osteoporosis, angiogenesis and/or hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoiesis systems, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues and cell types (e.g., endothelial cells, blood cells, bone, and gall bladder, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of bone formation and growth, osteonecrosis, osteoporosis, and/or hematopoiesis because its involvement in the growth signaling or angiogenesis.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 45**

The translation product of this gene shares sequence homology with signal sequence receptor gamma subunit which is thought to be important in protein translocation on endoplasmic reticulum.

This gene is expressed primarily in adrenal gland, salivary gland, prostate, and to a lesser extent in endothelial cells and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: protein secretion. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the secretory organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, salivary gland, prostate, endothelial cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to SSR gamma subunit indicate that polynucleotides and polypeptides corresponding to this gene are useful for endocrine disorders, prostate cancer, xerostomia or sialorrhea.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 46**

This gene is expressed primarily in osteoclastoma cells and to a lesser extent in melanocyte, amygdala, brain, and stomach.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ossification, osteoporosis, fracture, osteonecrosis, osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., melanocytes, amygdala, brain and other tissue of the nervous system, and stomach, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in intervention of ossification, osteoporosis, fracture, osteonecrosis and osteosarcoma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 48

The translation product of this gene shares sequence homology with proline rich proteins which is thought to be important in protein-protein interaction.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological and psychological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system and endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful in intervention

and detection of neurological diseases, including trauma, neoplasia, degenerative or metabolic conditions in the central nerve system. Additionally, the gene product may be secreted by the brain as an endocrine.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene shares sequence homology with the AOCB gene from *Aspergillus nidulans* which is important in asexual development.

This gene is expressed primarily in infant brain and to a lesser extent in the developing embryo, trachea tumors, B-cell lymphoma and synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative diseases, leukemia and sarcoma's. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, blood cells, trachea, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in infant brain and sarcoma's and homology to a gene involved in a key step of eukaryotic development (fungal spore formation) indicates that the protein product of this clone could play a role in neurological diseases such as schizophrenia, particularly in infants. The existence of the gene in a B-cell lymphoma indicates the gene may be used in the treatment and detection of leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary disorders including lung cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene only in fetal lung indicates that it plays a key role in development of the pulmonary system. This would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung. It may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 51

This gene is expressed primarily in hematopoietic cell types and fetal cells and to a lesser extent in all tissue types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in the immune system and hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene predominantly in hematopoietic cells and in the developing embryo indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of lymphomas and disease states affecting the immune system or hematopoiesis disorders such as leukemia, AIDS, arthritis and asthma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene is expressed primarily in prostate and to a lesser extent in fetal spleen, fetal liver, infant brain and T cell leukemias.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate disorders, prostate cancer, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, and/or prostate gland expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, spleen, liver, brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in prostate indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection or treatment of prostate disorders or prostate cancer. Its distribution in fetal liver and fetal spleen indicates it may play a role in the immune system and its misregulation could lead to immune disorders such as leukemia, arthritis and asthma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 53

The translation product of this gene shares sequence homology with dynein.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuro-degenerative diseases of the brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly neuro-degenerative diseases expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The predominant tissue distribution in the brain and homology to dynein, a microtubule motor protein involved in the positioning of cellular organelles and molecules indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection/treatment of neurodegenerative diseases, such as Alzheimers, Huntingtons, Parkinsons diseases and shizophrenia.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 54

The translation product of this gene shares sequence homology with ubiquitin-conjugation protein, an enzyme which is thought to be important in the processing of the Huntingtons Disease causing gene.

10 This gene is expressed primarily in brain and to a lesser extent in activated macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disease states including Huntington's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of brain tissues. For a number of disorders of the above tissues or cells, particularly of the neurological systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The predominant tissue distribution of this gene in the brain and its homology to a Huntington interacting protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the regulation of the expression of the Huntington disease gene and other neurodegenerative diseases including spinocerebellar ataxia types I and III, dentatorubropallidolysian and spinal bulbar muscular atrophy. In addition, the existence of elevated levels of free ubiquitin pools in Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis indicates that the ubiquitin pathway of protein degradation plays a role in these disease states. Thus, considering the gene described here is homologous to a ubiquitin-conjugation protein it may play a general role in neurodegenerative conditions.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 56

This gene is expressed primarily in T-cells (anergic T-cells, resting T-Cells, apoptotic T-cells) and lymph node (breast), as well as brain (hypothalamus, hippocampus, pituitary, infant brain, early-stage brain).

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune (e.g. immunodeficiencies, autoimmunities, inflammation, leukemias & lymphomas) and neurological (e.g. Alzheimer's disease, dementia, schizophrenia) disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous, hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, lymphoid tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the intervention or detection of pathologies associated with the hematopoietic and immune systems, such as anemias (leukemias). In addition, the expression in brain (including fetal) might suggest a role in developmental brain defects, neuro-degenerative diseases or behavioral abnormalities (e.g. schizophrenia, Alzheimer's, dementia, depression, etc.).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in lung, and to a lesser extent in a variety of other hematological cell types (e.g. Raji cells, bone marrow cell line, activated monocytes).

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary and/or hematological dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculo-pulmonary and hematopoietic systems, expression of this



39

- gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

- corresponding to this gene are useful in the intervention and detection of pathologies associated with the vasculo-pulmonary system. In addition the expression of this gene in a variety of leukocytic cell types and a bone marrow cell line might suggest a role in hematopoietic and immune system disorders, such as leukemias & lymphomas, inflammation, immunodeficiencies and autoimmunities.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 58

- 15 The translation product of this gene shares sequence homology with adenylyl kinase isozyme 3 (gil163528 GTP:AMP phosphotransferase (EC 2.7.4.10) [Bos taurus]), which is thought to be important in catalyzing the phosphorylation of AMP to ADP in the presence of ATP or inorganic triphosphate.

- 20 This gene is expressed primarily in fetal liver, heart and placenta, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as

- reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic, cardiovascular or reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic, cardiovascular and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, heart, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

- 35 corresponding to this gene are useful for the treatment and diagnosis of conditions related to hepatic function and pathogenesis, in particular, those dealing with liver development and the differentiation of hepatocyte progenitor cells.

40

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in CD34 positive cells (Cord Blood).

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic differentiation and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the detection and treatment of conditions associated with CD34-positive cells, and therefore as a marker for cell differentiation in hematopoiesis, as well as immunological disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 60

- 25 The translation product of the predicted open reading frame of this contig has sequence identity to the murine gene designated Insulin-Like Growth Factor-Binding Protein (IGFBP)-1 as described by Lee and colleagues (Hepatology 19 (3), 656-665 (1994)).

This gene is expressed exclusively in hemangiopericytoma.

- 30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hemangiopericytoma and other pericyte or endothelial cell proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and immune systems, expression of this gene at significantly higher or lower levels may routinely be detected in certain tissues and cell types (e.g., pericyte or endothelial cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Polynucleotides and polypeptides corresponding to this gene are useful as cell growth regulators since IGFBP-1-like molecules function as modulators of insulin-like growth factor activity. In addition, since IGFBP-1 is expressed at high levels following hepatectomy and during fetal liver development, polynucleotides of the present invention may also be used for the diagnosis of developmental disorders. Further, polypeptides of the present invention may be used therapeutically to treat developmental liver disorders as well as to regulate hepatocyte and supporting cell growth following hepatectomy or to treat liver disorders.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hemangiopericytoma and liver disorders.

15

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 61

This gene is expressed primarily in schizophrenic frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: nervous system and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the frontal cortex and CNS expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of frontal cortex, neuro-degenerative and CNS disorders

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 62

This gene is expressed primarily in human adrenal gland tumor, and to a lesser extent in human kidney, medulla and adult pulmonary tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic, endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and nervous system disorders and neoplasia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, kidney, brain and other tissue of the nervous system, pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of neurological and endocrine disorders including neoplasia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 63

20 This gene is expressed primarily in human adipocytes, and to a lesser extent in spleen, 12-week old human, and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune, metabolic and growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes, spleen, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of immune, developmental and metabolic disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 64**

One translated product of this clone is homologous to the mouse zinc finger protein PZF. (See Accession No. 453376; see also Gene 152 (2), 233-238 (1995).) Preferred polypeptide fragments correspond to the highly conserved domains shared between mouse and man. For example, preferred polypeptide fragments comprise the amino acid sequence: LQCEICGFTCRQKASLNWHMKKHADASFYQFSCNICGKFEKDSVVAHKAKSH PEV (SEQ ID NO: 621); ITSTDILQTNPESLTOPSD (SEQ ID NO: 622); NSTSGECLLEAEGM SKSY (SEQ ID NO: 623); CSGTERVSLMADGKIFVGSQSGQTGELVMNSDILGATTEVLIEDSD SAGP (SEQ ID NO: 624); IQYVRCMEGCGTVLAHPRYLQHHIKYQHLLKKKYVCPHPSCGRLF RLQKQLRHAKHHT (SEQ ID NO: 625); DQRDYICEYCARAFKSSHNLAVHRMIHTGEK (SEQ ID NO: 626); RSSRTSVSRHRDTEHTSSRSKTGSLQICKSEPTDQLDY (SEQ ID NO: 627); PFKDDPDETYKPHLERETPKRRKSG (SEQ ID NO: 630); QYVRCMEGCGTVLAHPRYLQ HHIKYQHLLKKKYVCPHPSCGRLFRLQKQLRHAKHHTD (SEQ ID NO: 629); or residues 151-182 of QRDYICEYCARAFKSSHNLAVHRMIHTGEKH (SEQ ID NO: 628). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Rhabdomyosarcoma, melanocyte and colon cancer tissue and to a lesser extent in smooth muscle, pancreatic tumor, and apoptotic T-cells.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hemopoietic, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striated muscle, melanocytes, colon, smooth muscle, pancreas, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of cancer and hemopoietic disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 65**

This gene is expressed primarily in human adipose and salivary gland tissue and to a lesser extent in human bone marrow and fetal kidney.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose, lower levels may be routinely detected in certain tissues and wounded tissues) or salivary gland, bone marrow, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis of metabolic and immune disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 66**

- 20 This translated product of this gene was recently identified as oxytocinase splice variant 1. (See Accession Nos. 2209276 and d1010078.) Preferred polypeptide fragments comprise the amino acid sequence: EMFDSLSTYFKSSLLMLKTYLSEDFVQHAVVLYLHN HSYASIQSDDLWDSFNEVTNQLDYKRMKKTWTLQKQFPLVTVQKKGKELFIQGERFPLNMK PEIQSDTRRYM (SEQ ID NO: 631). Also preferred are polynucleotide fragments encoding this polypeptide fragment.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 67**

- 30 This gene is expressed primarily in hemopoietic cells, particularly apoptotic T-cells, and to lesser extent in primary dendritic cells and adipose tissue.
- 35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of apoptotic T-cells, primary dendritic cells, and adipose tissue present in a biological sample and for diagnosis of diseases and conditions: hemopoietic diseases including cancer and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the oral and intestinal mucosa as well as hemopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment of diseases of the immune system, including cancer, hemopoietic and infectious diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 68

This gene is expressed primarily in kidney cortex and to a lesser extent in infant brain, heart, uterus, and blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of kidney tissue present in a biological sample and for diagnosis of diseases and conditions: soft tissue cancer, inflammation, kidney fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and endocrines systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, brain, and other nervous tissue, heart, uterus, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and fibroses.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 69

The translation product of this gene shares strong sequence homology with vertebrate and invertebrate protein tyrosine phosphatases.

This gene is expressed primarily in endometrial tumors, melanocytes, myeloid progenitors and to a lesser extent in infant brain, adipocytes, and several hematopoietic stem cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of transformed hematopoietic and epithelial cells present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of skin and endometrium, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, melanocytes, bone marrow, adipocytes, hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and sequence similarity with tyrosine phosphatases

indicate that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and hematopoietic disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 70

This gene is expressed primarily in osteoclastoma, breast, and infant brain and to a lesser extent in various fetal and transformed bone, ovarian, and neuronal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: degenerative conditions of the brain and skeleton. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of degenerative, neurological and skeletal disorders.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene was originally cloned from tumor cell lines. Recently another group has also cloned this gene, calling it the human malignant melanoma metastasis-suppressor (KISS-1) gene. (See Accession No. U43527.) Preferred polypeptide fragments comprise the amino acid sequence: LEKVASVGNSRPTGQQLSGLLA (SEQ ID NO: 632); VHRREASCYQQAEPGDL (SEQ ID NO: 633); RPALRQAGGTRPRQRKRYAQL (SEQ ID NO: 634); and AVNFRPQRSQSM (SEQ ID NO: 635). Any frame shifts can easily be resolved using known molecular biology techniques.

This gene is expressed primarily in many types of carcinomas and to a lesser extent in many normal organs.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanomas, and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of transformed organ tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. As a tumor suppressor gene, increase amounts of the polypeptide can be used to treat patients having a particular cancer.

25 The tissue distribution indicates that this gene and the translated product is useful for diagnosing and study of cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene is expressed primarily in striatum and to a lesser extent in adipocytes and hemangiopericytoma.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of striatal cells present in a biological sample and for diagnosis of diseases and conditions: neurological, fat and lysosomal storage

diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striatal tissue, adipocytes, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, study and treatment of neurodegenerative and growth disorders.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 73

This gene is expressed primarily in bone marrow stromal cells and to a lesser extent in smooth muscle, testes, endothelium, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of bone marrow present in a biological sample and for diagnosis of diseases and conditions: connective tissue and hematopoietic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, stromal cells, smooth muscle, testes and other reproductive tissue, endothelium, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of connective tissue and blood diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 74**

This gene is expressed primarily in brain, fetal liver and lung and to a lesser extent in retina, spinal chord, activated T-cells and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of brain and regenerating liver present in a biological sample and for diagnosis of diseases and conditions: CNS and spinal chord injuries, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, pulmonary tissue, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of hematopoietic and neurological conditions.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 75**

The translation product of this gene shares sequence homology with GTP binding proteins (intracellular).

This gene is expressed primarily in bone marrow, brain, and melanocytes and to a lesser extent in various endocrine and hematopoietic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic and nervous system conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, melanocytes, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to nucleotide binding factors indicate that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of brain degenerative, skin and blood diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 76**

This gene is expressed primarily in activated T-cells and to a lesser extent in retina, brain, and fetal bone.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of activated T-cells and developing brain present in a biological sample and for diagnosis of diseases and conditions: immune deficiencies and skeletal and neuronal growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and skeletomuscular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, brain and other tissue of the nervous system, retinal tissue, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, study and treatment of cancer, urogenital, and brain degenerative diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 77**

This gene is expressed primarily in fetal liver, activated monocytes, osteoblasts and to a lesser extent in synovial, brain, and lymphoid tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of myeloid and lymphoid present in a biological sample and for diagnosis of diseases and conditions: inflammation, immune deficiencies, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and skeleton, expression of this gene at significantly

higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, blood cells, bone, synovial tissue, brain and other tissue of the nervous system, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of lymphoid and mesenchymal cancers and nervous system diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 78

The translation product of this gene shares sequence homology with polymerase polypeptide precursor which is thought to be important in DNA repair and replication

This gene is expressed primarily in infant brain and to a lesser extent in tumors and tumor cell lines

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, especially of the neural system and developing organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to polymerase polypeptide precursor indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers especially of the neural system and developing organs

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 79

This gene is expressed primarily in muscle and endothelial cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., muscle, endothelial cells, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment and diagnosis of disorders of the vascular and neural system including cardiovascular and endothelial.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 80

This gene is expressed primarily in placenta and to a lesser extent in fetal liver

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental disorders and disorder of the haemopoietic system, fetal liver and placenta. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developmental disorders and disorder of the haemopoietic system, fetal liver and placenta, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of developmental disorders and disorders of the haemopoietic system, fetal liver and placenta.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 81**

This gene is expressed primarily in bone marrow, placenta and tissues and organs of the hematopoietic system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the bone and haematopoietic system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, bone and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, placenta, and hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the immune, bone and hematopoietic system

**FEATURES OF PROTEIN ENCODED BY GENE NO: 82**

The translation product of this gene shares sequence homology with secretory carrier membrane protein which is thought to be important in protein transport and export. Any frame shifts in coding sequence can be easily resolved using standard molecular biology techniques. Another group recently cloned this gene, calling it SCAMP. (See Accession No. 2232243.)

This gene is expressed primarily in prostate, breast and spleen, and to a lesser extent in several other tissues and organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the breast prostate and spleen. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly disorders of the breast prostate and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell

types (e.g., prostate, mammary tissue, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to secretory carrier membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the breast, prostate and spleen.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 83**

This gene is expressed primarily in developing organs and tissue like placenta and infant brain and to a lesser extent in developed organs and tissue like cerebellum and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, heart, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the neural system including neurological disorders and cancer.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 84**

The translation product of this gene shares sequence homology with ATPase 6 in Trypanosoma brucei which is thought to be important in metabolism.

This gene is expressed primarily in tumor and fetal tissues and to a lesser extent in melanocytes, kidney cortex, monocytes and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a



biological sample and for diagnosis of diseases and conditions: metabolism disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, melanocytes, kidney, blood cells, ovary and other tissue of the reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATPase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of metabolism disorders, especially in fetal and tumor tissue growth.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 85

The translation product of this gene shares sequence homology with the immunoglobulin superfamily of proteins which are known to be important in immune response and immunity.

20 This gene is expressed primarily in stromal cells, colon cancer, lung, amygdala, melanocyte and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells, colon, lung, amygdala, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution and homology to immunoglobulin indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune system disorders.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 86

The translation product of this gene shares sequence homology with transcription initiation factor eIF-4 gamma which is thought to be important in gene transcription.

This gene is expressed primarily in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis.

10 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in tumor tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution and homology to transcription initiation factor eIF-4 gamma indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene regulation in tumorigenesis.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 87

The translation product of this gene shares sequence homology at low level in prolines with secreted basic proline-rich peptide II-2 which is thought to be important in protein structure or inhibiting hydroxyapatite formation in vitro.

This gene is expressed primarily in endometrial tumor and fetal lung.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: endometrial tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular/skeletal and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample

taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to secreted basic proline-rich peptide II-2 indicate that polynucleotides and polypeptides corresponding to this gene are useful for inhibiting hydroxyapatite formation or establishing cell/tissue structure.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 88

This gene is expressed primarily in: amniotic cells induced with TNF in culture; and to a lesser extent in colon tissue from a patient with Crohn's Disease; parathyroid tumor; activated T-cells; cells of the human Caco-2 cell line; adenocarcinoma; colon; corpus colosum; fetal kidney; pancreas tumor; fetal brain; early stage brain, and anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system; e.g., tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., amniotic cells, colon, kidney, pancreas, parathyroid, brain and other tissue of the nervous system, blood cells, hematopoietic cells, liver, spleen, bone, testes and other reproductive tissue, brain and other tissue of the nervous system, and epithelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for modulating tumorigenesis and other immune system conditions such as disorders in immune response.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 89

This gene is expressed primarily in fetal liver/spleen and hematopoietic cells and to a lesser extent in brain, osteosarcoma, and testis tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, liver, spleen, bone, testes, and other reproductive tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 90

The translation product of this gene shares weak sequence homology with mouse Gcapi protein which is developmentally regulated in brain.

This gene is expressed primarily in infant and adult brain and fetal liver/spleen and to a lesser extent in smooth muscle, T cells, and a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, hematopoietic, immune, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, spleen, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and its homology to Ccapi1 protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosis of disorders in neuronal, hematopoietic, immune, and endocrine systems.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in brain and hematopoietic cells and to a lesser extent in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorder in nervous, hematopoietic, immune systems and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the in nervous, hematopoietic, immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of disorders in the nervous, hematopoietic, and immune systems.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 92

The translation product of this gene shares sequence homology with neuroendocrine-specific protein A which is thought to be important in neurologic systems.

This gene is expressed primarily in brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neural disorders and degeneration disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central or peripheral nervous systems, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to neuroendocrine-specific protein A indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neural disorders and degeneration disease.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 93

The translation product of this gene shares sequence homology with collagen-like protein and prolin-rich protein which are thought to be important in connective tissue function and tissue structure.

This gene is expressed primarily in fetal liver/spleen and brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to collagen-like protein and proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for supporting brain and hematopoietic tissue function and diagnosis and treatment of disorders in these functions.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 94

This gene is expressed primarily in embryonic tissues and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to,. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system (e.g., tumors), expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in brain tumor, placenta, and melanoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor or melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain or melanocytes, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, placenta, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the translation product of this gene is useful in the diagnosis and treatment of brain tumors and melanoma.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 96

The translation product of this gene shares sequence homology with a yeast membrane protein, SUR4, which encodes for APA1 that acts on a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.

This gene is expressed primarily in fetal liver; and to a lesser extent in placenta and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of fetal liver or defects of glucose-regulated ATPase activities in tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune/hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, placenta, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to yeast SUR4 membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of defects of fetal liver or defects of glucose-regulated ATPase activities.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 97

This gene is expressed primarily in fetal liver, brain, and amniotic fluid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the fetal immune system and adult brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune system and adult brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the fetal immune and hematopoietic systems since fetal liver is

the predominant organ responsible for hematopoiesis in the fetus. In addition, the gene product of this gene is thought to be useful for detecting certain neurological defects of the brain.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

The translation product of this gene shares sequence homology with an yolk protein precursor, Vitellogenin which is thought to be important in binding lipids such as phospholipids.

This gene is expressed primarily in amniotic cells and fetal liver.

- 10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in amniotic cells, fetal liver development and the fetal immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 25 The tissue distribution and homology to vitellogenin indicate that the protein product of this clone is useful for treatment and diagnosis of defects in amniotic cells, fetal liver development and the fetal immune system.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 99

- 30 This gene is expressed primarily in placenta, endometrial tumor, osteosarcoma and stromal cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumor of the endometrium or bone, and osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the obstetric system (e.g., placenta,

- endometrium) and the bones, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endometrium, bone, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors and abnormalities of the endometrium, and the bones because of its abundance in the aforementioned tissues..

## FEATURES OF PROTEIN ENCODED BY GENE NO: 100

This gene is expressed primarily in hepatocellular tumor.

- 15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatocellular tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 25 The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of hepatocellular cancer because of its abundant expression in this tissue.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 101

- 30 This gene is expressed primarily in Corpus Colosum, fetal lung and infant brain.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the Corpus Colosum or defects of the fetal lung. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Corpus Colosum and brain in general, and fetal lung, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of the Corpus Colosum and brain in general, and defects of fetal lung.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 102

This gene is expressed primarily in T cells and stromal cells, and to a lesser extent in adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of T cell immunity and stromal cell development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, stromal cells, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of T cell immunity and stromal cell development because of its abundant expression in these tissues.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 103

This gene is expressed primarily in infant brain and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, especially brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the brain, especially in young children.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 105

This gene is expressed primarily in human osteoclastoma and to a lesser extent in human pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly osteoclastoma and pancreatic tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in transformed tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone and pancreas, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of some types of tumors, particularly pancreatic cancer and osteoclastoma.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 106**

This gene is expressed primarily in fetal liver/spleen, and to a lesser extent in activated T-Cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of immune disorders.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 107**

This gene is expressed primarily in human embryo and to a lesser extent in spleen and chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the diagnosis and treatment of leukemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 108**

This gene is expressed primarily in placenta, and to a lesser extent in early stage human brain and in lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: fetal developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in fetal and amniotic tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this is useful for production of growth factor(s) associated with fetal development. Preferred polypeptides comprise the full-length polypeptide shown in the sequence listing, truncated however, at the amino terminus and beginning with QTTE.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 109**

This gene is expressed primarily in fetal spleen, and to a lesser extent in B-Cell lymphoma and T-Cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the treatment and diagnosis of human lymphomas.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 110

5 The translation product of this gene shares sequence homology with sarcoma amplified sequence (SAS), a tetraspan receptor which is thought to be important in malignant fibrous histiocytoma and liposarcoma.

This gene is expressed primarily in human osteoclastoma, and to a lesser extent in pineal gland and infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: malignant fibrous histiocytoma and liposarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, pineal gland, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution and homology to sarcoma amplified sequence (SAS) indicate that the protein product of this clone is useful for treatment of, osteosarcoma, malignant fibrous histiocytoma and liposarcoma and related cancers, particularly sarcomas.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 111

The translation product of this gene shares sequence homology with 6.8K proteolipid protein, mitochondrial - bovine.

30 This gene is expressed primarily in Wilm's tumor and to a lesser extent in cerebellum and placenta.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Wilm's tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the immune or renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution and homology to 6.8K proteolipid protein indicate that the protein product of this clone is useful for diagnostic and therapeutics associated with tumors, particularly Wilm's tumor disease.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 112

15 This gene is expressed primarily in embryonic tissue and to a lesser extent in osteoblasts, endothelial cells, macrophages (GM-CSF treated), and bone marrow.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, bone, endothelial cells, blood cells and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune disorders. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: MITDVQLAIFANMLGVSLFLVLYHYVA VNNPKKQE (SEQ ID NO: 636).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 113

35 This gene is expressed primarily in hepatocellular tumor, and to a lesser extent in fetal liver/spleen.



Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors, particularly hepatocellular tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of tumors, particularly hepatocellular tumors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene exhibits a very high degree of sequence identity with the human Pig8 gene which is thought to be important in p53 mediated apoptosis. The sequence of this gene has since been published by Polyak and colleagues (Nature 389, 300-306 (1997)). In addition, the predicted translation product of this contig exhibits very high sequence homology with a murine gene denoted as El24 which is also thought to be important in p53 mediated apoptosis.

This gene is expressed primarily in infant brain and activated T-cells and to a lesser extent in bone marrow, fetal liver, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and tissue damage by radiation and anti-cancer drugs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, bone marrow, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human Pig8 and murine El24 genes indicate that polynucleotides and polypeptides corresponding to this gene are useful for preventing apoptosis in patients being treated with anti-oncogenic drugs such as etoposide, hydroperoxycyclophosphamide, and X-irradiation, since this protein product is upregulated in cells undergoing such treatment where p53 was overexpressed. It may also be useful in the treatment of hematopoietic disorders and in boosting numbers of hematopoietic stem cells by interfering with the apoptosis of progenitor cells. The mature polypeptide is predicted to comprise the following amino acid sequence:

EEMADSVKTFQLDARGIKDSIWQICTSKLDARIQKREQRRASSVLAQRRASIERKQES  
EPRIVSRQCCAVNCGVWFSSLLFYRVFIVLQSVTARIGDPSLHQDVWSWLEFLLTSIFSA  
LWVLPFLVLSKVYNAVFOADIADLAFESVSGKKHPFSPVSKIADMLFNLLQLALFLQGMFVSL  
FPIHLVGQLVSLHMSLLVSLYCFEYRWNKGIEMHQRLSNERNWPYYFGFGLPLAFLTAMQ  
SSYISGCLFSLLPFLISANEAKTPGKAYLQRLRSLVFLSNRLEFKTVYLLQSALSSSTSAEK  
FSPHPSPAKLKATAGH (SEQ ID NO: 637). Accordingly, polypeptides comprising the foregoing amino acid sequence are provided as are polynucleotides encoded such polypeptides.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 115

This gene is expressed primarily in stromal cells and to a lesser extent in multiple sclerosis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: affecting the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of multiple sclerosis and other autoimmune diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 116**

This gene is expressed primarily in the gall bladder

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: gall stones or infection of the digestive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for possible prevention of digestive disorders where there may be a lack of digestive enzymes produced or in the detection and possible prevention of gall stones.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 117**

The translation product of this gene shares sequence homology with dystrophin gene which is thought to be important in building and maintenance of muscles.

This gene is expressed primarily in placenta and to a lesser extent in fetal brain and fetal liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: muscular dystrophy, Duchenne and Becker's muscular dystrophies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal muscle system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, muscle, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the dystrophin gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diseases related the degenerative myopathies that are characterized by the weakness and atrophy of muscles without neural degradation, such as Duchenne and Becker's muscular dystrophies.

**10 FEATURES OF PROTEIN ENCODED BY GENE NO: 118**

This gene is expressed primarily in olfactory tissue and to a lesser extent in cartilage.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: connective tissue diseases; chondrosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the connective tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue and cartilage, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for tumors of connective tissues, osteoarthritis and the treatment and diagnosis of chondrosarcoma.

**30 FEATURES OF PROTEIN ENCODED BY GENE NO: 119**

This gene is expressed primarily in Activated Neutrophils and to a lesser extent in fetal spleen, and CD34 positive cells from cord blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergies, defects in hematopoiesis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoiesis system the, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for reducing the allergic effects felt by allergy suffers by neutralizing the activity of the immune system, especially since neutrophils are abundant in persons suffering from allergies and other inflammatory conditions.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 120

The translation product of this gene shares sequence homology with poly A binding protein II which is thought to be important in RNA binding for transcription of RNA to DNA

This gene is expressed primarily in colon and to a lesser extent in brain and immune system.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: colon cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon, tissue and cells of the immune system, and brain or other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution and homology to poly A binding protein II indicate that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of colon cancer and other disorders of the digestive system..

#### 1 FEATURES OF PROTEIN ENCODED BY GENE NO: 121

The translation product of this gene shares sequence homology with thymidine diphosphoglucose 4,6 dehydrase which is thought to be important in the metabolism of sugar.

5 This gene is expressed primarily in fetal liver and spleen and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diabetes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution and homology to thymidine diphosphoglucose 4,6 dehydrase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of persons with diabetes since it appears that this protein is needed in the metabolism of sugar in to its more basic components.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 122

The translation product of this gene shares sequence homology with ceruloplasmin which is thought to be important in the metabolism and transport of iron and copper. Ceruloplasmin also contains domains with homology to clotting factors V and VIII. Defects in the circulating levels of ceruloplasmin (aceruloplasminemia) have been associated with certain disease conditions such as Wilson disease, and the accompanying hepatolenticular degeneration.

30 This gene is expressed primarily in brain and retina and to a lesser extent in endothelial cells.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases marked by defects in iron metabolism; aceruloplasminemia not characterized by defects in the

known ceruloplasmin gene locus; nonclassical Wilson disease; movement disorders; and tumors derived from a brain tissue origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, retina, and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, retinal tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ceruloplasmin indicate that

polynucleotides and polypeptides corresponding to this gene are useful for treatment of patients with aceruloplasminemia, or other defects in iron and/or copper metabolism. Mutations in this locus could also be diagnostic for patients currently experiencing or predicted to experience aceruloplasminemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 123

This gene is expressed primarily in brain and B cell lymphoma and to a lesser extent in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: B cell lymphoma; tumors and diseases of the brain and/or spleen; hematopoietic defects. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders in neuronal,

hematopoietic, and immune systems. It could potentially be useful for neurodegenerative disorders and neuronal and/or hematopoietic cell survival or proliferation.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in osteoclastoma, dermatofibrosarcoma, and B cell lymphoma and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer in particular osteoclastoma, dermatofibrosarcoma, and B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, immune, and circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, epidermis, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers and lymphoma; osteoporosis; and the control of cell proliferation and/or differentiation.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in immune tissues and hematopoietic cells, particularly in activated T cells and neutrophils, spleen, and fetal liver, and to a lesser extent in infant adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in T cell activation; hematopoietic disorders; tumors of a hematopoietic and/or adrenal gland origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and/or endocrine systems, expression of this gene at significantly higher

or lower levels may be routinely detected in certain tissues and cell types (e.g., cells and tissues of the immune system, hematopoietic cells, blood cells, liver, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune and/or hematopoietic disorders; diseases related to proliferation and/or differentiation of hematopoietic cells; defects in T cell and neutrophil activation and responsiveness; and endocrine and/or metabolic disorders, particularly of early childhood.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in placenta and endothelial cells and to a lesser extent in melanocytes and embryonic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell origin; angiogenesis associated with tumor development and metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, melanocytes, and embryonic tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of developmental disorders; inhibition of angiogenesis; and vascular patterning.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 127

This gene is expressed primarily in endothelial cells and hematopoietic tissues, including spleen, tonsils, leukocytes, and both B- and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell and/or hematopoietic origin; leukemias and lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, hematopoietic cells, spleen, tonsils, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the manipulation of angiogenesis; the differentiation and morphogenesis of endothelial cells; the proliferation and/or differentiation of hematopoietic cells; and the commitment of hematopoietic cells to distinct cell lineages.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 128

This gene is expressed primarily in kidney medulla and to a lesser extent in spleen from chronic myelogenous leukemia patients, prostate cancer, and some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a kidney origin; chronic myelogenous leukemia; prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, spleen, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of kidney disorders and cancer, particularly chronic myelogenous leukemia and prostate cancer. It may also be useful for the enhancement of kidney tubule regeneration in the treatment of acute renal failure.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 129

This gene is expressed primarily in adult and infant brain and to a lesser extent in mesenchymal or fibroblast cells, as well as tissues with a mesenchymal origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and of mesenchymal cells and tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; and fibrosis, based upon the expression of this gene within those tissues. Fibrosis is considered as mesenchymal cells and fibroblasts are the primary cellular targets involved in this pathological condition.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 130

This gene is expressed primarily in hepatocellular cancer and to a lesser extent in fetal tissues as well as testes tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, fetal tissue, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 131

This gene is expressed only in infant early brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: development and diseases of the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the brain in children and in treating nervous system disorders such as Alzheimer's disease, schizophrenia, dementia, depression, etc.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in brain and to a lesser extent in glioblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Alzheimer's disease,

schizophrenia, depression, mania, and dementia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating brain disorders such as Alzheimer's disease, schizophrenia, depression, mania, and dementia.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 133

The translation product of this gene shares sequence homology with ribitol dehydrogenase of bacteria which is thought to be important in metabolism of sugars.

This gene is expressed primarily in macrophage and to a lesser extent in T-cell lymphoma and lung.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tissue destruction in inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution and homology to ribitol dehydrogenase indicate that polynucleotides and polypeptides corresponding to this gene are useful for altering macrophage metabolism in diseases such as inflammation where macrophages are causing excess tissue destruction.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 134

This gene is expressed primarily in pancreatic tumor and to a lesser extent in synovial sarcoma.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, . Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and connective tissue systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pancreas, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing various cancers.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 135

This gene is expressed primarily in T cell lines such as Raji and to a lesser extent in infant brain.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune system disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing inflammatory diseases

such as rheumatoid arthritis, sepsis, inflammatory bowel disease, and psoriasis, as well as neutropenia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 136

The translation product of this gene shares high sequence homology with SAR1 subfamily of GTP-binding proteins which is thought to be important in vesicular transport in mammalian cells.

This gene is expressed primarily in serum-stimulated smooth muscle cells and to a lesser extent in a T-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases affecting vesicular transport. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP-binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene therapy in treating the large number of diseases involved in defective vesicular transport within cells..

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 137

The translation product of this gene shares sequence homology with a protein found in *C. elegans* cosmid F25B5.

This gene is expressed primarily in a fetal tissues and to a lesser extent in melanocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development, especially of the pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal pulmonary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, pulmonary tissue, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment and diagnosis of diseases affecting the pulmonary system, such as emphysema.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 138

This gene is expressed primarily in gall bladder and to a lesser extent in smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: digestive system disease and gall bladder problems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the digestive system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 139

This gene is expressed primarily in placenta and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development. Similarly, polypeptides and antibodies directed to these polypeptides are



useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developing tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing abnormal fetal development.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 140

This gene is expressed primarily in smooth muscle and to a lesser extent in ovary, prostate cancer, and activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hypertension and atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,

particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, ovary and other reproductive tissue, prostate, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the circulatory system, such as hypertension, atherosclerosis, etc.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 141

This gene is expressed primarily in fetal spleen and to a lesser extent in placenta and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and other diseases affecting blood cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen, placenta, bone marrow, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the generation of red and white blood cells and for the diagnosis of disease of these cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 142

The predicted translation product of this contig is a human homolog of the murine tetracycline/sugar transporter molecule recently reported by Matsuo and colleagues (Biochem. Biophys. Res. Commun. 238 (1), 126-129 (1997)).

This gene is expressed primarily in synovium and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: rheumatoid arthritis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and lymphatic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory diseases, such as rheumatoid arthritis, leukemia, neutropenia, inflammatory bowel disease, psoriasis, sepsis, and the like.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 143

This gene is expressed primarily in placenta and to a lesser extent in melanocyte, fetal liver and spleen, and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal early development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, melanocytes, liver, spleen, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of abnormal early development phenomena and diseases.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and neutropenia.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and blood systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the

35

expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in hematopoiesis and bone marrow regeneration as it is most abundant in fetal tissues responsible for the generation of hematopoietic cells.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 145

The translation product of this gene shares sequence homology with protein tyrosine phosphatase which is thought to be important in transducing signal to activate cells such as T cell, B cell and other cell types.

10

This gene is expressed primarily in T cells and tissues in early stages of development and to a lesser extent in cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-related diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic and fetal tissue, undifferentiated cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15

20

25

The tissue distribution and homology to the protein tyrosine phosphatase family indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating the immune system.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 146

This gene is expressed primarily in T cell and to a lesser extent in B cell, macrophages and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

35

providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

10 corresponding to this gene are useful for regulating the immune system therefore can be used in treating diseases such as autoimmune diseases and cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 147

This gene is expressed primarily in placenta and to a lesser extent in endothelial cells, testis tumor, ovarian cancer, uterine cancer.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, testis and ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 148

This sequence has significant homology to mouse torsin A. Recently, another group cloned the human Torsin A gene. (See, Accession No. Z358279; see also Nature Genet. 17, 40-48 (1997).)

35 This gene is expressed primarily in osteoclastoma, T-cell, and placenta and to a lesser extent in fetal lung, fetal liver, fetal brain, adult brain and tumor tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions; disease conditions in hematopoiesis and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, bone, placenta, lung, liver, and brain and other tissues of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating blood related diseases such as deficiencies in red blood cell, white blood cell, platelet and other hematopoiesis cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 149

20 This gene is expressed primarily in T cell, prostate and prostate cancer, endothelial cells and to a lesser extent in monocyte, dendritic cell, bone marrow, salivary gland, colon cancer, stomach cancer, pancreatic tumor, uterine cancer, fetal spleen and osteoclastoma.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions; immuno-related diseases and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, prostate, endothelial cells, dendritic cells, bone marrow, salivary gland, colon, stomach, pancreas, uterus, spleen and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 150

5 This gene was recently cloned by another group, calling it cIF3-p66. (See Accession No. 2351378.) This gene plays a role in RNA binding and macromolecular assembly, and therefore, any mutations in this gene would likely result in a diseased phenotype. Preferred polypeptide fragments comprise the amino acid sequence:  
MAKFMTPVIQDNPSQWGPCAVPEQFRDMPYQPFSGKDRLGKVDWTGATYQDKRYTNKYSS  
10 QFGGGSQYA YFHEDESSFLVDARTQKTAYQRNRMFAQRNLRDRRNMQLQFNQLQLP  
KSAKQKERIRLQKFKQGVGRQKWDQKSKQPRDSSVEVRSDWEVKEEMDFPQLMKMRY  
LEVSEPDIECCGALEY YDKAFDRITRSEKPLRXXKRUFHTVTTTDDPVIRKLAKTQGNVFATD  
AILATLMSCTRSVYSWDIVVQVSGSLFFDKRDNSDFDLLTVSETANEPPQDEGNSFNSPRL  
AMEATYINHNSQQCLRMGKERYNFPNPFVEDDDMKNEIASVAYRYSCKLGDDIDLIVRC  
15 EHDGVMTGANGEVFINIKTLNEWSRHCNGVDWRQLDSQGA VIATELKNNSYKLARWTC  
CALLAGSEYLKLGYYRYHVHVDSSRHVILGTQQFKPNEFASQINTSVENAWGILRCVIDICMKL  
EEGKYLIKDPNKQVIRVYSLPDGTFSS (SEQ ID NO: 638), as well as N-terminal and C-terminal deletions of this polypeptide fragment.

This gene is expressed primarily in T cell, bone marrow, embryo and endothelial cells and to a lesser extent in testis tumor and endometrial tumor.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders and cancers.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 151

This gene is expressed primarily in testis and to a lesser extent in T cell, spinal cord, placenta, neutrophil and monocyte.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: male reproductive and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive, immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testis and other reproductive tissue, blood cells, tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell  
10 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating immune and reproductive functions.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 152

The translation product of this gene shares sequence homology with tyrosyl-tRNA synthetase which is thought to be important in cell growth.

25 This gene is expressed primarily in brain, liver, keratinocytes, tonsils, and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, keratinocytes, tonsils, heart  
30 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissues of the nervous system, liver, keratinocytes, tonsils and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard

gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to tyrosyl-tRNA synthetase indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 153

This gene is homologous to the *Drosophila* transcriptional regulator *dre4*. (See Accession No. 2511745.) *Dre4* is a gene required for steroidogenesis in *Drosophila* melanogaster and encodes a developmentally expressed homologue of the yeast transcriptional regulator CDC68. Preferred polypeptide fragments comprise the amino acid sequence: KRRHTDVQFTTEVGETTDLGKHQMHDRDLVAEQMEREMRHKLTAFKN FLEKVEALTKBELFEFVFRDLGNQAPYRSTCLQPTSSALVNATEWPPVVTLDEVELHFEXR VQFHLKNPDMVIVYKDYSKKVTMAINAPVASLDPRIKEWLNQCDLKYTEGVQSLNWTMKMTVD DPEGFEEQGGWSFL (SEQ ID NO: 639), as well as N-terminal and C-terminal deletions of this fragments. Also preferred are polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in fetal liver, spleen, placenta, lung, T cell, thyroid, testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor, heart and liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal liver, spleen, placenta, lung, T cell, thyroid, testes expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, placenta, lung, blood cells, thyroid, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed primarily in brain and to a lesser extent in fetal heart, testis, spleen, lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: heart, liver and spleen diseases, immunological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, fetal heart, testis, spleen, lung expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, heart, testes and other reproductive tissue, spleen, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 155

Activation of T cells through the T cell antigen receptor (TCR) results in the rapid tyrosine phosphorylation of a number of cellular proteins, one of the earliest being a 100 kDa protein. This gene is the human equivalent of murine valosin containing protein (VCP). VCP is a member of a family of ATP binding, homo-oligomeric proteins, and the mammalian homolog of *Saccharomyces cerevisiae* cdc48p, a protein essential to the completion of mitosis in yeast. Both endogenous and expressed murine VCP are tyrosine phosphorylated in response to T cell activation. Thus we have identified a novel component of the TCR mediated tyrosine kinase activation pathway that may provide a link between TCR activation and cell cycle control.

This gene is expressed primarily in brain, liver, spleen, placenta. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, spleen, placenta expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, spleen, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to VCR indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 156

The translation product of this gene shares sequence homology with rat growth response protein which is thought to be important in cell growth. A group recently cloned the human homolog of this gene, calling it insulin induced protein 1. (See Accession No. 2358269, see also, Genomics 43 (3), 278-284 (1997).) Preferred polypeptide fragments comprise the amino acid sequence: RSGLGICITATLITQF L V Y N G V Y Q T S P D F L Y I R S W L P C I F F S G G V T V G N I G R Q L A M G V P E K P H S D (SEQ ID NO: 640). as well as N-terminal and C-terminal deletions of this polypeptide fragment. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain, liver, placenta, heart, spleen, lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, placenta, heart, spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, placenta, heart, spleen, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to growth-response protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 157

This gene is expressed primarily in Glioblastoma, endometrial tumor, lymphoma and pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Glioblastoma, Endometrial tumor, lymphoma and pancreas tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, lymphoid tissue, pancreas, and tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 158

The translation product of this gene shares sequence homology with ICE receptor which is thought to be important in allergy and asthma.

This gene is expressed primarily in T cell, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergy and asthma and other immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to IgE receptor indicate that polynucleotides and polypeptides corresponding to this gene are useful for allergy and asthma.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 159

The translation product of this gene shares sequence homology with immunoglobulin heavy chain which is thought to be important in immune response to the antigen.

This gene is expressed primarily in activated neutrophil and to a lesser extent in activated T cell, monocyte and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: infection, inflammation and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to immunoglobulin heavy chain variable region indicate that polynucleotides and polypeptides corresponding to this gene are useful for making the ligand to block specific antigen which cause certain disease.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 160

The translation product of this gene shares sequence homology with mouse X inactive specific transcript protein which is thought to be important in X chromosome inactivation.

This gene is expressed primarily in HSA172 cell and to a lesser extent in normal ovary tissue, ovarian cancer, frontal cortex and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ovarian tumor, schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to X inactive specific transcript protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive system tumors and CNS tumors.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 161

This gene is expressed primarily in adipose cell and to a lesser extent in liver and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity and liver disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose cell, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose cells, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of obesity and liver disorder.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 162

The translation product of this gene shares sequence homology with yeast ubiquitin activating enzyme homolog which is thought to be important in protein posttranslational processing.

This gene is expressed primarily in stromal cell and to a lesser extent in retina, H. Atrophic Endometrium, colon carcinoma and myeloid progenitor cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development, neuronal growth disorders and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal cells, endometrium, colon, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ubiquitin-activating enzyme homolog indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of some type of tumors, fucosidosis and neuronal growth disorders.

20

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 163

This gene is expressed primarily in primary breast cancer and hemangiopericytoma and to a lesser extent in adult brain and cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer, leukemia and cerebellum disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of various tumors and disease involved in neural system.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 164

The translation product of this gene shares sequence homology with proline rich proteins. Recently, another group has also cloned this gene, calling it CD84 leukocyte antigen, a new member of the Ig superfamily. (See Accession No. U82988, see also, Blood 90 (6), 2398-2405 (1997).)

10 This gene is expressed primarily in Weizmann olfactory tissue and osteoclastoma and to a lesser extent in anergic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ostsis and immune disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue, bone, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution and homology to the Ig superfamily indicate that the protein product of this clone is useful for treatment of osteoporosis, autoimmune disease, and other immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 165

30 This gene is expressed primarily in atrophic endometrium and colon cancer and to a lesser extent in some fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system,

35



expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, colon, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors, specifically endometrium and colon tumors.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human primary breast cancer and to a lesser extent in activated monocyte. Although the predicted signal sequence is identified in Table 1, other upstream sequences are also relevant. Preferred polypeptide fragments comprise the amino acid sequence: VTQPKHLASMGVPEIPSFYYPWELAXXPXVRISWRGHFHG QSFYSTRPSIHKDYVNRLELNMTEQGESQFLRISNLKEDQSVYFCRVELDTRRSQ (SEQ ID NO: 641), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of breast cancer.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 167

This gene is expressed primarily in fetal tissues and to a lesser extent in adult lung. This gene has also been mapped to chromosomal location 9q34, and thus, can be used as a marker for linkage analysis for chromosome 9.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryo tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 168

The translation product of this gene shares sequence homology with Ig Heavy Chain which is thought to be important in immune response.

This gene is expressed primarily in prostate cancer tissue specifically. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The translation product of this gene shares sequence homology with cytosolic acyl coenzyme-A hydrolase, which is thought to be important in neuron-specific fatty acid metabolism. The gene represented by this contig has since been published by Hajra and colleagues (GenBank Accession No. U91316).

This gene is expressed primarily in human pituitary gland and to a lesser extent in colorectal cancer tissue. This gene has also been observed in the LNCAP cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hyperlipidemias of familial and/or idiopathic origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly blood, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to rat cytosolic acyl coenzyme-A hydrolase indicate that polynucleotides and polypeptides corresponding to this gene are useful for the detection or treatment of hyperlipidemia disease states by virtue of the ability of specific drugs to activate the enzyme.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 170

The translation product of this gene shares sequence homology with a *Caenorhabditis elegans* gene which is thought to be important in organism development.

This gene is expressed primarily in human synovial sarcoma tissue, bone marrow, and to a lesser extent in human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of bone, specifically synovial sarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, connective tissues and possibly immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, bone marrow, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another

tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to *Caenorhabditis elegans* indicate that polynucleotides and polypeptides corresponding to this gene are useful as a diagnostic and/or therapeutic modality directed at the detection and/or treatment of connective tissue sarcomas or other related bone diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 171

The translation product of this gene shares sequence homology with beta1-6GlcNAc transferase which is thought to be important in the transfer and metabolism of beta1-6, N-acetylglucosamine. This gene product has previously been shown to suppress melanoma lung metastasis in both syngeneic and nude mice, decreased invasiveness into the matrigel, and inhibition of cell attachment to collagen and laminin without affecting cell growth.

This gene is expressed primarily in human testes and prostate tissues, and to a lesser extent in kidney, medulla, and pancreas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, prostate, kidney, pancreas, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to beta1-6GlcNAc transferase indicate that the protein product of this clone is useful for the development of diagnostic and/or therapeutic modalities directed at the detection and/or treatment of cancer, the metastasis of malignant tissue or cells. Defects in this potentially secreted enzyme may play a role in metastasis.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 172**

This gene is expressed primarily in fetal spleen and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders, Wilms tumor disease, hepatic disorders, and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and identification of fetal defects along with correcting diseases that affect hematopoiesis and the immune system.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 173**

The translation product of this gene shares sequence homology with ret II oncogene which is thought to be important in Hirschsprung disease and many types of cancers.

25 This gene is expressed in multiple tissues including the lymphatic system, brain, and thyroid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Hirschsprung disease and multiple cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue, thyroid, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to ret II oncogene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of various cancers. It would also be useful for the diagnosis and treatment of Hirschsprung disease. Preferred polypeptides of the invention comprise the amino acid sequence: MEAQVNEAESAREQLQXHDQAQKASKQELTELRLKQEFHYIEEDLY RTKNTLQSRUKDRDEIEIKLRNQLTNKTLSNSQSELENRLHQLTETLIQKQTMLESISTEKNLS VFQLERLQQMNSASQSSNGSSINMSCIDNCGETLRRVPLFNDTETNLACMYCKYRKAAS SIDQSRILGFLRRYPLARVFVIYMLLHLVVMIVLLTTPEMHHDPYQK (SEQ ID NO: 642).

**10 FEATURES OF PROTEIN ENCODED BY GENE NO: 174**

The translation product of this gene shares sequence homology with testis enhanced gene transcript which is thought to be important in regulation of human development.

15 This gene is expressed primarily in infant brain and to a lesser extent in a variety of other tissues and cell types, including the prostate, testes, monocytes, macrophages, dendritic cells, keratinocytes, and adipocytes.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological, developmental, immune and inflammation disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, prostate, testes and other reproductive tissue, blood cells, keratinocytes, and adipocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution and homology to testis enhanced gene transcript indicate that the protein product of this clone is useful for diagnosis and treatment of disorders involving the developing brain and the immune system.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 175**

This gene is expressed primarily in prostate and to a lesser extent in various other tissues, including placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers, especially of the prostate. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of prostate disorders and cancer. It may also be useful for the diagnosis and treatment of endocrine disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 176**

The translation product of this gene shares sequence homology with *Saccharomyces cerevisiae* YNT20 gene which is thought to be important in mitochondrial function.

This gene is expressed at a particularly high level in muscle tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases related to such tissues and cell types including: muscle wasting diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., muscle and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the YNT20 gene indicate that this protein is useful for treatment and detection of neuromuscular diseases caused by loss of mitochondrial function. For example this gene or its protein product could be used in replacement therapy for such diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 177**

This gene is expressed primarily in the brain and to a lesser extent in kidney, placenta, smooth muscle, heart and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuromuscular diseases, degenerative diseases of the central nervous system, and heart disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, central nervous system, and heart, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, kidney, placenta, muscle, heart and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

This gene or its protein product could also be used for replacement therapy for the above mentioned diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 178**

The translation product of this gene shares sequence homology with caldesmon which is thought to be important in the cellular response to changes in glucose levels.

This gene is expressed primarily in multiple tissues including brain and retina.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: central nervous system disorders and retinopathy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the CNS disorders and retinopathy, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and retinal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to caldesmon indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of retinopathies.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 179

The translation product of this gene shares sequence homology with mouse fibrosin protein which is thought to be important in regulation of fibrogenesis in certain chronic inflammatory diseases.

This gene is expressed primarily in amniotic cells and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of breast cancer and abnormal embryo

development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to fibrosin indicate that the protein product of this clone is useful for treatment of breast cancer. This gene or its protein product could be used in replacement therapy for breast cancer. In addition the protein product of this gene is useful in the treatment of chronic inflammatory diseases.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 180

This gene is expressed several infant tissues including brain and liver and various adult tissues including brain, lung, liver, testes, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, brain cancer, lung cancer, liver cancer and cancers of the reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, hepatic system, and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, liver, testes and other reproductive tissue, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene product indicates that the protein product of this clone is involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 181

This gene is expressed primarily in activated monocytes and to a lesser extent in melanocytes and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of immune system diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, melanocytes, and dendritic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 182**

This gene is expressed primarily in placenta and several tumors of various tissue origin and to a lesser extent in normal tissues including liver, lung, brain, and skin,

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers of all kinds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

10 of the above tissues or cells, particularly of the central nervous system, respiratory system and skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, lung, brain and other tissues of the nervous system, and skin, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

15 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The high expression of this gene in multiple tumors indicates that the protein product of the clone may be involved in cell growth control and therefore would be useful for treatment of certain cancers. Likewise molecules developed to block the

20 activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 183**

25 The translation product of this gene shares sequence homology with the mouse Ndr1 gene which is thought to be important in cancer progression.

This gene is expressed multiple cell types and tissues including brain, lung, kidney, bone marrow, liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as

30 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and endocrine

35 systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, kidney, bone marrow, liver and spleen, and cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to Ndr1 gene, which is thought to be involved in cancer progression, indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of certain cancers. Likewise molecules developed to block the activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 184**

10 This gene is expressed primarily in early stage human brain and liver and to a lesser extent in several other fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain and liver cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

15 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue

20 or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 185**

This gene is expressed primarily in infant and embryonic brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of degenerative nervous system disorders and brain cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful

35 in providing immunological probes for differential identification of the tissue(s) or cell

115

type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 186

This gene is expressed primarily in multiple tissues including placenta, fetal lung, fetal liver, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers including liver, brain and lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, pulmonary system, and hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, lung, liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

116

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	11	582	1	582	177	177	313	1	18	19	22
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	197	1020	296	830	442	442	499	1	18	19	22
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	12	465	1	465	81	81	314	1	30	31	128
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	198	524	229	343		196	500	1	20	21	33
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	13	474	1	474	1	1	315	1	24	25	28
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	199	332	1	319	35	35	501	1	24	25	28

[illegible]



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	23	1486	596	1418	102	102	325	1	54	55	252
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	203	847	1	839	87	87	505	1	30	31	75
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	204	852	75	850		690	506	1			10
13	HTXEF04	209235 09/04/97	Uni-ZAP XR	205	1354	54	1354	100	100	507	1	33	34	207
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	24	2323	1017	2059	1242	1242	326	1	21	22	68
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	206	1378	113	1226	303	303	508	1	25	26	36
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	25	683	1	683	304	304	327	1	30	31	84

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	207	1166	281	884	567	567	509	1	18	19	19
16	HHFFL33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	26	2036	14	1959	214	214	328	1	20	21	36
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	27	717	1	717	70	70	329	1	30	31	63
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	208	697	2	697	33	33	510	1	31	32	32
18	HMDAE90	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	28	495	1	495	39	39	330	1	24	25	35
19	HOUAW01	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	29	556	1	556	116	116	331	1	19	20	23



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
28	HSSDM73	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	38	672	1	672	22	22	340	1	38	39	42
29	HBMVK68	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	39	1908	135	1908	309	309	341	1	20	21	26
30	HMKDC66	97898 02/26/97 209044 05/15/97	pSport1	40	458	93	458	147	147	342	1	24	25	26
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	41	1153	500	1153	427	427	343	1	30	31	157
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	213	1079	502	896		739	515	1	23	24	43
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	42	1983	1092	1983	27	27	344	1	11	12	520

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	214	3791	2757	3357		2030	516	1			3
33	HTOJN06	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	43	1406	1	695		19	345	1	19	20	39
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	44	1391	851	1153	74	74	346	1	30	31	234
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	215	1334	822	1036		638	517	1	18	19	174
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	45	1569	768	1569	14	14	347	1	19	20	169
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	216	1511	770	1404	844	844	518	1	32	33	43



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
44	HMABL38	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	220	1258	149	1190	254	254	522	1	18	19	26
45	HSKDK47	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	55	1896	596	1614	650	650	357	1	33	34	47
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	56	1753	555	1753	414	414	358	1	18	19	73
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	221	1693	554	1693		526	523	1	25	26	58
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	57	1220	690	1024	128	128	359	1	30	31	102
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	222	1196	712	1163		1097	524	1			19

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
48	HFCAl74	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	58	1049	362	1049	335	335	360	1	33	34	48
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	59	1776	854	1737	189	189	361	1	30	31	179
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	223	1791	979	1791	1164	1164	525	1	18	19	40
50	HLFBC91	97899 02/26/97 209045 05/15/97	pBluescript SK-	60	443	1	443	164	164	362	1	21	22	25
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	61	2888	1909	2888	90	90	363	1	30	31	224
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	224	2517	1597	2517	1953	1953	526	1	18	19	57



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
59	HCWEF90	97899 02/26/97 209045 05/15/97	ZAP Express	230	448	9	448		1	532	1	22	23	75
60	HHGCM20	97899 02/26/97 209045 05/15/97	Lambda ZAP II	70	245	1	245	93	93	372	1	1	2	51
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	71	361	1	361	1	1	373	1	30	31	61
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	231	407	1	407	210	210	533	1	17	18	60
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	72	713	8	713	169	169	374	1	30	31	40
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	232	830	190	580	329	329	534	1	28	29	39

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	73	862	1	862	67	67	375	1	30	31	44
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	233	932	138	905	287	287	535	1			2
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	74	4602	4162	4525	730	730	376	1	30	31	203
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	234	2786	2406	2739	2577	2577	536	1	22	23	36
65	HSGBA84	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	75	1255	1	1195	112	112	377	1	28	29	29
66	HTOCD52	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	76	475	1	475	13	13	378	1	30	31	136

Gene No.	cDNA Clone ID	ATCC Deposit No.: Z and Date	Vector	NT SEQ ID NO:	Total NT Seq.	5' NT B. NT	5' NT of Clone	Seq.	Codon Start	5' NT	AA First of AA of ID of Y	AA First of AA of ID of Y	AA First of AA of ID of Y	AA First of AA of ID of Y	AA First of AA of ID of Y	AA First of AA of ID of Y	AA First of AA of ID of Y
71	HFFAH94	97900 02/26/97 209046 05/15/97	Lambda ZAP II	81	1290	768	1054	701	701	383	1	21	22	138			
72	HBIA195	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	82	684	1	684	119	119	384	1	30	31	74			
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	83	2024	1609	1953	200	200	385	1	30	31	521			
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	237	1286	391	959		1204	539	1	9	10	11			
74	HEBEG68	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	84	931	14	537	85	85	386	1	25	26	137			
75	HBIBAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	85	825	59	802	66	66	387	1	30	31	186			

[illegible]



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
75	HBIAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	238	734	1	734	1	1	540	1	37	38	108
75	HBIAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	239	809	80	794		294	541	1	15	16	106
76	HTXDU73	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	86	1238	36	918	17	17	388	1			1
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	87	1460	9	1458	166	166	389	1	53	54	299
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	240	2201	841	2080	507	507	542	1	43	44	136
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	241	1661	311	1520	390	390	543	1	35	36	424

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
78	HTEIY30	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	88	1395	567	1395	639	639	390	1	36	37	49
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	89	1186	352	1186	540	540	391	1	49	50	61
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	242	1146	329	1146	564	564	544	1	21	22	39
80	HPMFL27	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	90	1821	1203	1614	1503	1503	392	1	30	31	79
81	HMWDN32	97900 02/26/97 209046 05/15/97	Uni-Zap XR	91	862	253	862	359	359	393	1	32	33	36
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	92	696	349	696	98	98	394	1	30	31	180



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
91	HTSEL31	97901 02/26/97 209047 05/15/97	pBluescript	101	1394	608	1346	602	602	403	1	23	24	87
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	102	794	1	794	518	518	404	1	30	31	92
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	248	1766	42	1766	356	356	550	1	30	31	168
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	249	2664	47	1708		147	551	1	18	19	124
93	HODAS59	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	103	1544	898	1531	975	975	405	1			21
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	104	871	106	871	248	248	406	1	34	35	174

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	250	865	97	865	258	258	552	1	19	20	177
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	105	404	1	404	16	16	407	1	21	22	64
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	251	2082	852	2074	829	829	553	1	22	23	72
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	106	1542	506	1542	122	122	408	1	51	52	280
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	252	1482	508	1482		633	554	1	15	16	45
96	HCUHB01	209215 08/21/97	ZAP Express	253	834	1	834	82	82	555	1	40	41	251
97	HAQBF73	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	107	2327	1528	2327	465	465	409	1	30	31	284

[illegible][illegible]

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	114	1171	540	1171	337	337	416	1	30	31	163
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	261	1179	626	1161	335	335	563	1	30	31	253
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	262	1162	629	1131	942	942	564	1			18
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	115	842	373	800	100	100	417	1	65	66	174
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	263	735	290	735			565	1			
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	264	783	416	783		413	566	1	33	34	73

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	116	1640	187	1470	581	581	418	1	30	31	50
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	265	1638	301	1405	119	119	567	1	30	31	263
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	266	1455	148	1188	438	438	568	1	24	25	70
107	HE6DK18	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	117	952	418	906	499	499	419	1	28	29	120
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	118	1256	21	1079	301	301	420	1	30	31	159
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	267	1086	25	1050	227	227	569	1	23	24	34



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	126	431	1	431	73	73	428	1	38	39	47
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	273	515	1	515	43	43	575	1	20	21	30
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	127	3752	3465	3752	748	748	429	1	30	31	370
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	274	2995	2738	2995	2777	2777	576	1	18	19	29
118	HASAS24	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	128	1144	669	1144	896	896	430	1			30
119	HSIDN55	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	129	1830	1234	1830	1265	1265	431	1			24

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
120	HGBGZ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	130	1864	1505	1741	1578	1578	432	1	37	38	53
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	131	2041	1	1214	46	46	433	1	35	36	176
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	275	1990	8	1128	71	71	577	1	16	17	92
122	HOECP43	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	132	2012	853	1986	1127	1127	434	1	22	23	77
123	H2CBV31	97902 02/26/97 209048 05/15/97	pBluescript SK-	133	1669	670	1632	962	962	435	1	25	26	32
124	HPCAD23	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	134	1565	281	1565	274	274	436	1	25	26	30





Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	145	1021	526	1021	74	74	447	1	30	31	278
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	278	961	524	961	545	545	580	1	23	24	110
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	146	1285	5	1285	116	116	448	1	30	31	199
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	279	1228	9	1228	324	324	581	1	26	27	30
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	147	1386	169	1272	165	165	449	1	30	31	258
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	280	1327	169	1208	160	160	582	1	23	24	71

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
138	HMWGI25	97902 02/26/97 209048 05/15/97	Uni-Zap XR	148	2098	721	2044	784	784	450	1	18	19	87
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	149	1847	1689	1847	241	241	451	1	33	34	315
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	281	799	1	799		243	583	1	12	13	47
140	HMSKE75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	150	1569	113	1517	417	417	452	1	21	22	52
141	HCM SH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	151	1540	538	1540	48	48	453	1	30	31	383
141	HCM SH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	282	2196	270	2196	294	294	584	1	32	33	39



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
150	HSHCC16	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	160	2120	1223	2108	1416	1416	462	1			14
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	161	900	482	900	46	46	463	1	30	31	285
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	288	1517	783	1517	1062	1062	590	1			24
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	162	1003	1	1003	288	288	464	1	30	31	80
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	289	3865	217	1195	281	281	591	1	16	17	38
153	HTSFQ12	97903 02/26/97 209049 05/15/97	pBluescript	163	2196	1607	2180	1611	1611	465	1	30	31	47

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	164	1945	271	1840	299	299	466	1	63	64	96
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	290	1910	279	1818	355	355	592	1	39	40	69
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	165	2933	489	2871	258	258	467	1	30	31	399
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	291	3276	486	2838		525	593	1	45	46	308
156	HJPCJ76	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	166	2243	343	2221		341	468	1			1
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	167	1816	1130	1816	284	284	469	1	31	32	273

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total Seq.	5' NT	3' NT	5' NT	Start Codon	5' NT	First AA of SEQ	Y	Sig of AA	First AA	Secreted of AA	Last AA	ORF
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	292	1695	1098	1548	1306	1306	594	1						22
158	HMKBA64	97903 02/26/97 209049 05/15/97	pSport1	168	945	1	787	208	208	470	1	18		19			192
159	HNFI24	97903 02/26/97 209049 05/15/97	pBluescript	169	902	46	816	19	19	471	1	26	27				234
160	HCLEB21	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	170	1883	798	1869	1001	1001	472	1	45	46				105
160	HCLEB21	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	293	1501	438	1501	510	510	595	1		24				
161	HAWBA28	97903 02/26/97 209049 05/15/97	pBluescript SK-	171	2100	1642	2100	1722	1722	473	1	23	24				32

162	HSAAS44	97903 02/26/97 209049 05/15/97	pBluescript SK-	172	1930	187	1930	65	65	474	1	30	31				571
162	HSAAS44	97903 02/26/97 209049 05/15/97	pBluescript SK-	294	2683	183	2683	431	431	596	1		24				
163	HAFAL73	97903 02/26/97 209049 05/15/97	pBluescript SK-	173	1509	962	1451	122	122	475	1	30	312				
163	HAFAL73	97903 02/26/97 209049 05/15/97	pBluescript SK-	295	1454	961	1420	976	976	597	1		1				
164	HSAWF26	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	174	3173	2197	2972	51	51	476	1	21	22				329
164	HSAWF26	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	296	828	52	828	305	305	598	1		8				

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	175	991	374	970	60	60	477	1	24	25	178
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	297	2416	1387	2413	1473	1473	599	1	18	19	25
166	HFKFX55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	176	1290	499	1290		688	478	1	25	26	52
167	H2LAO11	97903 02/26/97 209049 05/15/97	pBluescript SK-	177	2290	1	2290	173	173	479	1	22	23	62
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	178	549	1	549	11	11	480	1	21	22	27
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	298	545	1	545	17	17	600	1	21	22	27

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	179	1509	294	1352	92	92	481	1	30	31	339
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	299	1530	385	1530	562	562	601	1	23	24	61
170	HCFAE79	97904 02/26/97 209050 05/15/97	pSport1	180	1316	985	1250	995	995	482	1	26	27	32
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	181	777	1	777	51	51	483	1	30	31	48
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	300	997	244	997	300	300	602	1	23	24	29
172	HODCW06	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	182	791	1	791	14	14	484	1	29	30	38



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	190	1014	703	1014	360	360	492	1	30	31	159
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	306	577	1	577		175	608	1			6
181	HAFUA18	97904 02/26/97 209050 05/15/97	pBluescript SK-	191	2779	2207	2630	1153	1153	493	1	30	31	279
181	HAFUA18	97904 02/26/97 209050 05/15/97	pBluescript SK-	307	2860	163	2860	21	21	609	1	30	31	232
181	HAFUA18	97904 02/26/97 209050 05/15/97	pBluescript SK-	308	876	275	876	302	302	610	1	32	33	34
182	HETBY74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	192	1923	30	1923	45	45	494	1	33	34	193

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	193	2346	1160	2286	178	178	495	1	30	31	205
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	309	2025	840	2025	971	971	611	1	18	19	21
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	194	3054	2004	3054	434	434	496	1	11	12	147
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	310	3026	1966	3026		2131	612	1			9
185	HEBGB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	195	907	152	907	297	297	497	1	30	31	64
185	HEBGB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	311	712	67	712	107	107	613	1	18	19	29

Gene No.	186	HFAMH74	02/26/97	05/15/97	209050	05/15/97
cDNA Clone ID	186	HFAMH74	02/26/97	05/15/97	209050	05/15/97
ATCC Deposit No: Z and Date	97904	97904	02/26/97	05/15/97	209050	05/15/97
NT SEQ ID NO: X	196	1289	312	312	312	312
NT SEQ ID NO: Y	196	1289	312	312	312	312
NT SEQ ID NO: Z	196	1289	312	312	312	312
NT SEQ ID NO: W	196	1289	312	312	312	312
NT SEQ ID NO: V	196	1289	312	312	312	312
NT SEQ ID NO: U	196	1289	312	312	312	312
NT SEQ ID NO: T	196	1289	312	312	312	312
NT SEQ ID NO: S	196	1289	312	312	312	312
NT SEQ ID NO: R	196	1289	312	312	312	312
NT SEQ ID NO: Q	196	1289	312	312	312	312
NT SEQ ID NO: P	196	1289	312	312	312	312
NT SEQ ID NO: O	196	1289	312	312	312	312
NT SEQ ID NO: N	196	1289	312	312	312	312
NT SEQ ID NO: M	196	1289	312	312	312	312
NT SEQ ID NO: L	196	1289	312	312	312	312
NT SEQ ID NO: K	196	1289	312	312	312	312
NT SEQ ID NO: J	196	1289	312	312	312	312
NT SEQ ID NO: I	196	1289	312	312	312	312
NT SEQ ID NO: H	196	1289	312	312	312	312
NT SEQ ID NO: G	196	1289	312	312	312	312
NT SEQ ID NO: F	196	1289	312	312	312	312
NT SEQ ID NO: E	196	1289	312	312	312	312
NT SEQ ID NO: D	196	1289	312	312	312	312
NT SEQ ID NO: C	196	1289	312	312	312	312
NT SEQ ID NO: B	196	1289	312	312	312	312
NT SEQ ID NO: A	196	1289	312	312	312	312

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO: X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO: X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No: Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO: X. The nucleotide position of SEQ ID NO: X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO: X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO: Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO: Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO: Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO: Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO: X and the translated SEQ ID NO: Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO: X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO: X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO: Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.



Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide

sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988).

Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

### Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeech, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heijne, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heijne, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeech and von Heijne are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

### 10 Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

"Identity" per se has an art-recognized meaning and can be calculated using published techniques. (See, e.g.: COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, (1988); BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, (1993); COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, (1994);

SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, (1987); and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, (1991).) While there exists a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans. (Carillo, H., and Lipton, D., SIAM J Applied Math 48:1073 (1988).) Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in "Guide to Huge Computers," Martin J. Bishop, ed., Academic Press, San Diego, (1994), and Carillo, H., and Lipton, D., SIAM J Applied Math 48:1073 (1988).

30 Methods for aligning polynucleotides or polypeptides are codified in computer programs, including the GCG program package (Devereux, J., et al., Nucleic Acids Research (1984) 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Aischul, S.F. et al., J. Molec. Biol. 215:403 (1990)), Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711 (using the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981)).

When using any of the sequence alignment programs to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters are set so that the percentage of identity is calculated over the full length of the reference polynucleotide and that gaps in identity of up to 5% of the total number of nucleotides in the reference polynucleotide are allowed.

5 A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brulag et al. (Comp. App. Biosci. 6:237-245 (1990).) The term "sequence" includes nucleotide and amino acid sequences. In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB search of a DNA sequence to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, and Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter. Preferred parameters employed to calculate percent identity and similarity of an amino acid alignment are: Matrix=PAM 150, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in amino acid residues, whichever is shorter.

20 As an illustration, a polynucleotide having a nucleotide sequence of at least 95% "identity" to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone, means that the polynucleotide is identical to a sequence contained in SEQ ID NO:X or the cDNA except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the total length (not just within a given 100 nucleotide stretch). In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to SEQ ID NO:X or the deposited clone, up to 5% of the nucleotides in the sequence contained in SEQ ID NO:X or the cDNA can be deleted, inserted, or substituted with other nucleotides. These changes may occur anywhere throughout the polynucleotide.

30 Further embodiments of the present invention include polynucleotides having at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the polynucleotides having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity

will encode a polypeptide identical to an amino acid sequence contained in SEQ ID NO: Y or the expressed protein produced by the deposited clone.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference polypeptide, is intended that the amino acid sequence of the polypeptide is identical to the reference polypeptide except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the total length of the reference polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence. Further embodiments of the present invention include polypeptides having at least 80% identity, more preferably at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to an amino acid sequence contained in SEQ ID NO: Y or the expressed protein produced by the deposited clone. Preferably, the above polypeptides should exhibit at least one biological activity of the protein.

In a preferred embodiment, polypeptides of the present invention include polypeptides having at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98%, or 99% similarity to an amino acid sequence contained in SEQ ID NO: Y or the expressed protein produced by the deposited clone.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an

organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level.

Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., *J. Biotechnology* 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (*J. Biol. Chem.* 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make

phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln; replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

### **Polynucleotide and Polypeptide Fragments**

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, and 701 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, and 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about"

includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

### Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1983)).

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., *Proc. Natl. Acad. Sci. USA* 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., *Cell* 37:767-778 (1984); Sutcliffe, J. G. et al., *Science* 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., *Proc. Natl. Acad. Sci. USA* 82:910-914; and Bittle, F. J. et al., *J. Gen. Virol.* 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., *J. Nucl. Med.* 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

### Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In

preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).) Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the claimed invention.

#### Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and lac promoters, the SV40 early and late promoters and promoters of retroviral LTRs; to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS,

293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and puc99a, pKK223-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pVLNEO, pSV2CAT, pOC44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

### Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycle. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage

analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the

present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erllich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.



In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20

millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S. W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S. W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention could be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

### Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules

may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

#### Immune Activity

5 A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders 10 may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

15 A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: 20 blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, DiGeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome,

25 lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria. Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a 30 polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks 35 (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from

inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders. 5

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune 10 inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility. 15

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD. 20

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g. septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or 25 IL-1.)

30

### Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

15 Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thyroid, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

20 Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

### Infectious Disease

30 A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

20 Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillus, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Brucella, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Hemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria,

Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis,

Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas.

These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### **Regeneration**

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

### **Chemotaxis**

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat

disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

### Binding Activity

5 A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

10 Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

15 Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

20 The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

25 Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

30 Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The

antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

5 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

### Other Activities

10 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

15 A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

20 A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, cardiac rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

25 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

### Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone. A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining

whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:X in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the

amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a

polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an

amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least



199

- 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

- Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

- Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

- Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

- Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated

200

polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

- Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

### Examples

#### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

- Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lambd BA	plasmid BA
pSport1	pSport1
PCMVVSport 2.0	PCMVVSport 2.0
PCMVVSport 3.0	PCMVVSport 3.0
pCR <sup>2.1</sup>	pCR <sup>2.1</sup>

- Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988), Ailing-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Ailing-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation

of the fl origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the fl ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR<sup>®</sup>2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., BioTechnology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported.

The oligonucleotide is labeled, for instance, with <sup>32</sup>P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection

agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 μl of reaction mixture with 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl<sub>2</sub>, 0.01% (w/v) gelatin, 20 μM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to

remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

#### **Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO.X, according to the method described in Example 1. (See also, Sambrook.)

#### **Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P<sup>32</sup> using the rediprimer™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPD-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

#### **Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of

conditions : 30 seconds, 95°C, 1 minute, 56°C, 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

#### **Example 5: Bacterial Expression of a Polypeptide**

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9, (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis. Clones containing the desired constructs are grown overnight (ON) in liquid culture in LB media supplemented with both Amp (100 µg/ml) and Kan (25 µg/ml). The ON culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.<sub>600</sub>) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic

agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising a phase operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number XXXXXX.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

#### Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a

stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Comassie blue stained 16% SDS-PAGE gel when 5  $\mu$ g of purified protein is loaded.

The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

#### Example 7: Cloning and Expression of a Polypeptide in a Baculovirus

##### Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAC373, pVL941, and pACIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as

required. Such vectors are described, for instance, in Luckow et al., Virology 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five  $\mu$ g of a plasmid containing the polynucleotide is co-transfected with 1.0  $\mu$ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One  $\mu$ g of BaculoGold™ virus DNA and 5  $\mu$ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50  $\mu$ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10  $\mu$ l Lipofectin plus 90  $\mu$ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to S9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200  $\mu$ l of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5  $\mu$ Ci of  $^{35}$ S-methionine and 5  $\mu$ Ci  $^{35}$ S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

#### 25 Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIV and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden),

pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., BioTechnology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the

naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100-200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

#### Example 9. Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having

more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No.209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGGCCAAATCTTCTGACAAAACATCACACATGCCACCGTGCC  
CAGCACTGAAATTCGAGGGTGCACCGTCAGTCTTCCCTTCCCCCAAACC  
CAAGGACACCCCTCATGATCTCCCGAGCTCTGAGGTACATGCGTGTGGT  
GGACGTAAGCCACGAAAGACCCTGAGGTCAAGTTCAACTGTGTACGTGACG  
GCGTGAGGAGTGATATGCCAAGACAAAGCCGCGGAGGAGACGATACAAC  
AGCACTAACCGTGTGTGACGCGTCTACCGTCTGCACGAGACTGGCTG  
AATGGCAAGGAGTACAAGTGCAAGGTCTCCACAAAGCCCTCCCAACCCCC  
ATCGAGAAAAACCATCTCCAAAGGCCAAAGGGACGCCGAGAAACACAGGT  
GTACACCCCTGCCCATCCCGGATGAGTGAACCAAGAACGAGTCAAGCT  
GACCTGCGCTGTCAAAAGGCTTCTATCCAAAGCGACATGCGCGTGAAGTGGGA  
GAGCAATGGGAGCGCGGAGAAACAACACTACAAGAACCAAGCCTCCCGTCTGG  
ACTCCGACGGCTCTTCTTCTCTCTACAGCAAGCTCACCGTGGACAAAGAGCA  
GGTGCCAGCAGGGGAACGCTTCTCATGCTCCGTGATGCATGAGGCTCTGC  
ACAACCACTACACGACGAAAGAGCTCTCCCTGTCTCCGGGTAATGAAGTGC  
GACGGCCGGAGCTTAGAGGAT (SEQ ID NO:1)

**Example 10: Production of an Antibody from a Polypeptide**

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody

whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')<sub>2</sub> and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

**Example 11: Production Of Secreted Protein For High-Throughput Screening Assays**

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhitaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10<sup>5</sup> cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhitaker))/10% heat inactivated FBS(14-503F Biowhitaker)/1x Penstrep(17-602E Biowhitaker). Let the cells grow overnight.



215

The next day, mix together in a sterile solution basin: 300  $\mu$ l Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2 $\mu$ g of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50 $\mu$ l of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150 $\mu$ l Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200 $\mu$ l of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstep, or CHO-5 media (see below) with 2mm glutamine and 1x penstep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50  $\mu$ l for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300 $\mu$ l multichannel pipetter, aliquot 600 $\mu$ l in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

216

### HGS-CHO-5 medium formulation:

#### Inorganic Salts

CaCl <sub>2</sub> (anhyd)	116.6 mg/L
CuSO <sub>4</sub> ·5H <sub>2</sub> O	0.00130
Fe(NO <sub>3</sub> ) <sub>3</sub> ·9H <sub>2</sub> O	0.050
FeSO <sub>4</sub> ·7H <sub>2</sub> O	0.417
KCl	311.80
MgCl <sub>2</sub>	28.64
MgSO <sub>4</sub>	48.84
NaCl	6995.50
NaHCO <sub>3</sub>	2400.0
NaH <sub>2</sub> PO <sub>4</sub> ·H <sub>2</sub> O	62.50
Na <sub>2</sub> HPO <sub>4</sub>	71.02
ZnSO <sub>4</sub> ·7H <sub>2</sub> O	.4320

#### Lipids

Arachidonic Acid	.002 mg/L
Cholesterol	1.022
DL-alpha-Tocopherol-Acetate	.070
Linoleic Acid	0.0520
Linolenic Acid	0.010
Myristic Acid	0.010
Oleic Acid	0.010
Palmitic Acid	0.010
Palmitic Acid	0.010
Pluronic F-68	100
Stearic Acid	0.010
Tween 80	2.20

#### Carbon Source

D-Glucose	4551 mg/L
-----------	-----------

#### Amino Acids

L-Alanine	130.85 mg/ml
L-Arginine-HCL	147.50
L-Asparagine-H <sub>2</sub> O	7.50
L-Aspartic Acid	6.65
L-Cystine-2HCL-H <sub>2</sub> O	29.56
L-Cystine-2HCL	31.29
L-Glutamic Acid	7.35
L-Glutamine	365.0
Glycine	18.75
L-Histidine-HCL	52.48

H <sub>2</sub> O	
L-Isoleucine	106.97
L-Leucine	111.45
L-Lysine HCL	163.75
L-Methionine	32.34
L-Phenylalanine	68.48
L-Proline	40.0
L-Serine	26.25
L-Threonine	101.05
L-Tryptophan	19.22
L-Tyrosine-2Na-2H <sub>2</sub> O	91.79
L-Valine	99.65

#### Vitamins

Biotin	0.0035 mg/L
D-Ca Pantothenate	3.24
Choline Chloride	11.78
Folic Acid	4.65
L-Inositol	15.60
Niacinamide	3.02
Pyridoxal HCL	3.00
Pyridoxine HCL	0.031
Riboflavin	0.319
Thiamine HCL	3.17
Thymidine	0.365
Vitamin B <sub>12</sub>	0.680

#### Other Components

HEPES Buffer	25 mM
Na Hypoxanthine	2.39 mg/L
Lipoic Acid	0.105
Sodium Putrescine-2HCL	0.081
Sodium Pyruvate	55.0
Sodium Selenite	0.0067
Ethanolamine	20uM
Ferric Citrate	0.122
Methyl-B-Cyclodextrin complexed with Linoleic Acid	41.70
Methyl-B-Cyclodextrin complexed with Oleic Acid	33.33
Methyl-B-Cyclodextrin complexed with Retinal Acetate	10

Adjust osmolality to 327 mOsm

#### Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

219

	ISRE Ligand	Tyk2	JAKs			STATs	GAS(element) or
			Jak1	Jak2	Jak3		
5	IFN family IFN- $\alpha$ /B IFN- $\gamma$ (IRF1>Ly6>IFP) IL-10	+	+	-	-	1,2,3 1 - 1,3	ISRE GAS
10	gpl30 family IL-6 (Pleiotrohic) (IRF1>Ly6>IFP) IL-11 (Pleiotrohic) OnM (Pleiotrohic) LIF (Pleiotrohic) CNTF (Pleiotrohic) G-CSF (Pleiotrohic) IL-12 (Pleiotrohic)	+	+	+	?	1,3 1,3 1,3 1,3 1,3 1,3 1,3 1,3	GAS
15	IL-2 (lymphocytes) IL-4 (lymph/myeloid) >Ly6 (d $\beta$ H) IL-7 (lymphocytes) IL-9 (lymphocytes) IL-13 (lymphocytes) IL-15	-	+	-	+	1,3,5 6 5 5 5 6 5	GAS GAS (IRF1 = IFP) GAS GAS GAS GAS GAS
20	gC family IL-3 (myeloid) (IRF1>IFP>Ly6) IL-5 (myeloid) GM-CSF (myeloid)	-	-	+	-	5 5 5 5	GAS GAS GAS GAS
25	gpl40 family IL-3 (myeloid) (IRF1>IFP>Ly6) IL-5 (myeloid) GM-CSF (myeloid)	-	-	+	-	5 5 5 5	GAS GAS GAS GAS
30	Growth hormone family GH PRL EPO CAS>IRF1=IFP>Ly6)	?	+	+	-	5 5 1,3,5 5	GAS GAS GAS GAS(B-
35	Receptor Tyrosine Kinases EGF	?	+	+	-	1,3	GAS (IRF1)
40	PDGF CSF-1	?	+	+	-	1,3 1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994)), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5'-GGCCCTCGAGATTTCGCCGAAATCTAGATTTCGCCGAAATGATTTCGCCG  
AAATGATTTCGCCGAAATATCTGCCATCTCAATTAG.3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5'-GGCGCAAGCTTTTGCAAAAGCCTAGGC.3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5'-CTCGAGATTTCGCCGAAATCTAGATTTCGCCGAAATGATTTCGCCGAAATG  
ATTTCGCCGAAATATCTGCCATCTCAATTAGTCAGACAACCATAGTCCGCC  
CTAACTCCGCCCATCCGCCCTAACTCCGCCAGTTCGCCCATTTCTCCGC  
CCCATGGCTGACTAATTTTATTATTATGACAGAGGCCGAGCCGCCCTCGGC  
CTCTGAGCTATTCCAGAGTAGTGAAGAGGCTTTTGGAGGCTTAGGCTTT  
TGCAAAAGCCT.3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS-SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a

neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

### Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jak-STAT signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI

+ 10% serum with 1% Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies) with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final concentration of  $10^7$  cells/ml. Then add 1ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat/GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at 20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 40°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

#### **Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Khatabanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^6$  U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting  $1 \times 10^6$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in the 96-well plate (or  $1 \times 10^5$  cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

#### **Example 15: High-Throughput Screening Assay Identifying Neuronal Activity**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1) (Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAAGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)  
5' GCGAAGCTTCGGGACTCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS/SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS/SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$  cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### **Example 16: High-Throughput Screening Assay for T-cell Activity**

NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I- $\kappa$ B is phosphorylated and degraded, causing NF- $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in treating diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those diseases related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTCCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:

5'-GCGGCTCGAGGGGACTTCCCGGGGACTTCCGGGACTTCCGGGAC.  
TTTCCATCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5'-GCGGCAAGCTTTTGC AAAAGCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2- (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5'-CTCGAGGGGACTTCCCGGGGACTTCCGGGAGACTTCCGGGACTTTCC  
ATCTGCCATCTCAATTAGTCAGCAACCAATAGTCCGCCCTAACTCCGCCCA  
TCCCGCCCTAACTCCGCCAGTTCCGCCCATTTCTCGCCCATGGCTGACT  
AATTTTATTATGACAGAGCGCGAGGCCCTCGGCTCTGAGCTATTTC  
CAGAAAGTAGTGAGGAGGCTTTTGGAGGCCCTAGGCTTTTGCAAAAAGCTT:  
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2- promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the

NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NcoI.

- Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

#### Example 17: Assay for SEAP Activity

- 10 As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

- 15 Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

- Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room

- 20 temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

- 25 Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

#### Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4

15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10<sup>6</sup> cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension.

The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10<sup>6</sup> cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular

signaling even which has resulted in an increase in the intracellular Ca<sup>++</sup> concentration.

**Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity**

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF, and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are



used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

- To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200µl/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 µl of the supernatant produced in Example 11, the medium was removed and 100 µl of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 40°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 µm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 40°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

- Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- The tyrosine kinase reaction is set up by adding the following components in order: First, add 10µl of 5µM Biotinylated Peptide, then 10µl ATP/Mg<sub>2</sub><sup>+</sup> (5mM ATP/50mM MgCl<sub>2</sub>), then 10µl of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTa, 100mM MgCl<sub>2</sub>, 5 mM MnCl<sub>2</sub>, 0.5 mg/ml BSA), then 5µl of Sodium Vanadate(1mM), and then 5µl of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10µl of the control enzyme or the filtered supernatant.
- The tyrosine kinase assay reaction is then terminated by adding 10 µl of 120mM EDTA and place the reactions on ice.

- Tyrosine kinase activity is determined by transferring 50 µl aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This

allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300µl/well of PBS four times. Next add 75 µl of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5µl/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

- Next add 100µl of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

#### **Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

- As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

- Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1µg/ml) for 2 hr at room temp. (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

- A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 µl of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

10

**Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products is then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals is identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenin deoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human col-1 DNA for specific hybridization to the corresponding genomic locus.

35

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera

(Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

10

**Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect soluble polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on

35

the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

### Example 23: Formulating a Polypeptide

5 The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

10 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1  $\mu\text{g/kg/day}$  to 10  $\text{mg/kg/day}$  of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01  $\text{mg/kg/day}$ , and most preferably for humans between about 0.01 and 1  $\text{mg/kg/day}$  for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1  $\mu\text{g/kg/hour}$  to about 50  $\mu\text{g/kg/hour}$ , either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

20 Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

30 The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include poly(lactides) (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Siddman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric

acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

10 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

15 Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

20 The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as poly(vinylpyrrolidone); amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium, and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

30 The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of

about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile.

5 Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

10 The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

#### Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

30 For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

#### Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

5 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

#### Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin, is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

25 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to

transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphiphotropic PA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is being produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

(1) GENERAL INFORMATION:

(i) APPLICANT: Human Genome Sciences, Inc. et al.

(ii) TITLE OF INVENTION: 186 Human Secreted Proteins

(iii) NUMBER OF SEQUENCES: 644

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.

(B) STREET: 9410 Key West Avenue

(C) CITY: Rockville

(D) STATE: Maryland

(E) COUNTRY: USA

(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage

(B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE: March 6, 1998

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: A. Anders Brookes, Esq.

(B) REGISTRATION NUMBER: 36.373

(C) REFERENCE/DOCKET NUMBER: PS002.PCT

## (2) INFORMATION FOR SEQ ID NO: 2:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser  
1 5

## (2) INFORMATION FOR SEQ ID NO: 3:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 86 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGCTCTGAG ATTTCCTGGA AATCTAGATT TCCCGAAT GATTTCCTCG AAATGATTTC  
60  
CCGGAATAT CTCCATCTC AATTAG  
86

## (2) INFORMATION FOR SEQ ID NO: 4:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGCTCTGAG ATTTCCTGGA AATCTAGATT TCCCGAAT GATTTCCTCG AAATGATTTC  
60  
CCGGAATAT CTCCATCTC AATTAG  
86

## (2) INFORMATION FOR SEQ ID NO: 5:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATT CCCGGAATC TAGATTCCC GGAATGATT TCCCGAAT GATTTCCTCG  
60

## (vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8439

## (2) INFORMATION FOR SEQ ID NO: 1:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 733 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GCGCAATCT TCTGACAAA CTCACATG CCCACCTGC CCAGCAGTG  
60  
AATTGAGGG TGACCGTCA GTCTTCTCT TCCGCCAAA ACCGAGGAC ACCCTCATGA  
120  
TCTTCCGGAC TCTGAGGTC ACATGCTGTG TGTGAGCTT AGCCAGCAA GACCTGAGG  
180  
TCAGTTTCAA CTGTAGCTG GACCGGTGG AGGTGCATTA TCCACAGCA AAGCCGGGG  
240  
AGGACAGTA CACAGCAGG TACGCTGTGG TCAGCTTCT CACCTTCTG CACAGGACT  
300  
GGCTGAATG CAGGAGTAC AAGTGCAGG TCTTCACAAA AGCCCTTCCA ACCCTCATG  
360  
AGAAACCTT CTCGAAAGC AAGGCGAGC CCGGAGACC ACAGGTATC ACCCTGCCC  
420  
CATCCCGGA TGACCTGACC AAGAACAGG TCAGCTTGC CTCCTGTGC AAGGCTTCT  
480  
ATCCAGGGA CATGCGGTG GAGTGGAGA GCAATGGGA GCGGAGAGC AACTACAGA  
540  
CCAGGCTTC CTTCTGTGAC TCGAGGCTT CTTTCTTCT CTACAGCAG CTCACCTGG  
600  
ACAGAGCAG GTGACAGCAG GGGAGCTCT TCTATGCTC CTGATGAT CATGAGCTCTGC  
660  
ACAACCACTA CAGCGAGAG AGCCTCTTCC TGTCTCCGG TAAATGATG CGAGGCGGC  
720  
GACTCTAGAG GAT  
733

AAATATCTGC CATCTCAATT ATCTAGACAC CATACCTCCG CCCCCTACTC CCCCACATCC 120  
 GCGCTTACTT CCGCCCAATT CCGCCCAATT TCCCCCCTAT GCGTACTTAA TTTTCTTTAT 180  
 TTTATCGAG CCGGAGCCCG CCGGAGCTTC TGAAGTATTC CAGAGATAGT GAGGAGCCTT 240  
 TTTTGAGGC CTACGCTTTT GCAAAAAGCT T 271

(2) INFORMATION FOR SEQ ID NO: 6:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 32 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(1X1) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGCTGAGG GATGAGAGCG ATTAGACCCC GG 32

(2) INFORMATION FOR SEQ ID NO: 7:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(1X1) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCGACATTC GCGATCCCG GATTCGCGCT C 31

(2) INFORMATION FOR SEQ ID NO: 8:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 12 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(1X1) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCGGACTTTC CC 12

(2) INFORMATION FOR SEQ ID NO: 9:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(1X1) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GCGGCTTGA GCGGACTTTC CCGGAGACTT TCCGGGAGCT TTCCGGAGCT TTCCATCTCG 60  
 CCACTCAAT TAG 73

(2) INFORMATION FOR SEQ ID NO: 10:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 256 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(1X1) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTCCTCCCG GACTTTCGCG GACTTTCGCG GACTTTCGCA TCTCCATCT 60  
 CATTATCTCA GCAACCTTAG TCCCTCCCTT AACCTCCGCC ATCCGAGCCC TAACTCCGCC 120  
 CAGTCCCGCC CATTCTCCCG CCGATGCGCTG ACTATTTTCT TTTATTTTATG CAGAGGCCGA 180  
 GCGGCGCTCG GCGTCTGAGC TATTCAGAA GTATGAGAGA GCGTCTTTTG GAGGCTTAGG 240  
 CTTTCCAAA AAGCTT 256

(2) INFORMATION FOR SEQ ID NO: 11:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 582 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(1X1) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GCGACAGAGT AATTTCTTAC AGAATTTCC AGACATTTAT GTACGTGAAA AAAATTCGA 60  
 GCAAGCTTAT AAGATCTTG GATCCATTA TATGATATGT ATAGCTGAAA TGTGTATTC 120  
 AATGACTTTT TCTCTTTAT CCTGAGCA AAAATTTGTT TAAATTTGCA TCCGAATGT 180  
 TTTTATCTT TGTATTTTCT TTAATAATCC TTTCTCTCT ATCATTTGCT TTTTCTGCT 240  
 TGTAAATAGA CTTACTTGCA CTTGGAAGAT GATTTACTCC TTGTATCTT ACAATATGT 300  
 GATATGCTAA TTTTCATTAAC AGATTCAGT TTTGAGCAA GATATGCTGA TTTGTTTATA 360  
 ACAAAAAC TCGCTCATT TCTGTGAAT TGTCTTTGA AAATTTCTTT TTACAGTGT 420

AGCCAACTG AGATACCTG ATGGCTTGA TTCTTTTCA TGTGCTTAC CATCTATTT 480  
AGCACTGAG CCTTTTATTA TTCTCTATT TGTAAAGTTT ATTGTCTTA ACTCATTTA 540  
TAATATTA CTATTACT GTTTATCTGT TTCTGAAAAA AAAAAAAAAA AA 582

10

(2) INFORMATION FOR SEQ ID NO: 12:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 465 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GTTTGGGGGT GAGGCCGAGC TGCTGGGGG CTTCGTGCGC GGCCAGACA CAGCTACTG 60  
CAGGCGGGG GCGCTTGCT ATGATGTC TCACCCAGGG GGGGCTCTG CCTCTACTC 120  
GTGCGAGCC CACTTGCCAG CGAGAGGCC TCCCAGGCC TTGAGGGCTG CTCGGAGTCA 180  
CCTGTGGAA TGGACTAAA GAGCCTTGT GTGGAGAG GTGCTGCCA AACAGCCTGC 240  
TCTGTGCTC CAGCCAGGCC CTCGTGAGG GAGGGGGCAG GACTCATCAG GACCTCCCTG 300  
GACCCCTGCA GGGCAGCAG CTGGGGCCG AGCCAGACA TTGTGGCTG CTGCCCCCA 360  
GGGACAGCA AGCCTCTTGG GCTCTTCCC TTCTGGACA AGGCCCCCTG CCTTTGCTC 420  
ACATAAACTG TACATATTT TCAATAAAG CCTCTTTCAT AAAA 465

40

(2) INFORMATION FOR SEQ ID NO: 13:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 474 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATGCATTTCC TGCTCAGAC CTTCCTGTG GTGCACTTC TGCTCTTTG TGTATGCC 60  
ATATGCTTAG GCTTCTCCCC CTCCTAGAG GGTCTCTTGA TAGATTAGA AATAAGATG 120  
AGTGACATTT CTAATGTGCA TATAGAGG AGCCAGACA CATGCTTTT AATTAAGG 180  
ACAGTGTCCA TCGTTTAGC TGCCGATAG AACCTTGGT TCATCTGCT GAGCTAGGC 240  
CTTTAAACA GCTTCTGTG TTCTCATTTG TCTAGTGT TTGCGAGGT TTATCGGA 300  
AGATATGTT CGTTTAAA TTTTCTTA TGAGGCCGG CGTGCTGCT CAGCCCTGA 360

ACCTAGCAG TTGGGGCTG AGCGGTGGA TCACGAGTC AGGAGATGCA GACCATCTG 420  
GSTAACATG TGAACCCCG TCTCTACTA AATACAAA AAAAAAAAAA AAA 474

10

(2) INFORMATION FOR SEQ ID NO: 14:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 314 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TTATGTTGG GAGCAAGACC TGATAGCCAG CCTTTACATG GGAGTATAT TCTGTCTCC 60  
ATCTCATAG CCCAGTACC TGAGCGAGA TGTATTATAC CAACCAACT GTCTCTTAT 120  
CATGATGGC TTTCAGATA GGTATTTC ATCATTTCCA TTGTAGCTC TACAGTGGT 180  
TATATTAAT TCTCATCTT TAAGTCTCT CCTCAGTCC TGTGTATTG AAATCATG 240  
CTCTCAGC CAGTTGAGCT CTGCATCTC CATTATGGG CAGAGCTGTG TTGAGAGAG 300  
AGATATNAG TTCC 314

35

(2) INFORMATION FOR SEQ ID NO: 15:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 613 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CTCATATTC CTTCTGGTA AAGTGACA TGCATGAT CATCTGCT TTAATTAT 60  
ATCTTCTTA TGTGCGAG CAGACAGA AGTAGAAGA AGATGCTGT AGCTCAGA 120  
ACCACTAAA TCTATCTAT GCGCTGGTT CACCCAGCT GCTTGTGGA TTTTGTCTA 180  
CTATAACAG GCTGCCAGG AGACTGAGA GTACGCTCC TTAAGCACTG TAACTAAGC 240  
CTAATCTTC CTTTCCACC AACATGTTC CGACTCATC CTCCTTCCR AGTCCCTTT 300  
TCTGCCCGC ATGCGATTG CATTTAATA ATCTCAGT GAATGTCCA CACAGATTC 360  
CATTTTAAT AGCATACCAT AGTTTTTGT CAATTTTCT TTCAGARGAC TCCATTGCA 420  
GCTGCTAGA GAGCTAAG GCGAGGCTC TTGAGCTTT CCCGATAGT TTCAGCTGCA 480  
ATAGCTCTTA GGCAGATGC CATGAGCTC CTGCCCCACT GTATTACTGG GGNACACTG 540



ATTGCGTAGA AGTTGATGCT CCGTGAACCT TTCTGAGCTC TTACATCTTA CTCGTGAAAC 600  
CGAAATATCC GAC 613

(2) INFORMATION FOR SEQ ID NO: 16:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 356 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CCCCCCCC TGAACCTGG GCTGTGAAG TTTTGGCTG TGTGTGCTG TGTGTGCTC 60  
GCTGTGCTG TGTGTGCTG CTCCTGTTC AAGTGTGAG CAGCCATCA TGAAGCCCC 120  
TTATTTTAC TTGCATGCA CCAAGTCTC CCCCCAGC CTCCTGTGAG TCCCTCATG 180  
GTATGTGCTC TCCCTGCGA AGAAGCTCA TTGTGTGAAA ATGCATCAT GTAAATTAAT 240  
GGAAAGCAT TTGTGCTGAG TTATGTTTAT TACATCATCA TTGCATCTG GAACTGCAT 300  
CCCTGAAGC GTAAATTTGAG GTGTACACA AGAAGCAGAG GGGAAAAA AAAAAA 356

(2) INFORMATION FOR SEQ ID NO: 17:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 414 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GAAATTAAT CCCCCGCTT TTAAAGCTTA CTCGTGAAT AAGTATGAG TAACTCATTA 60  
GGAGAGCTG AGTCAGACA ACGAAGCTA TAGAAGAAA TGAATGACT TTAAAAAAT 120  
ATTGACATTA ACTGTCTAG GAACTCTTC TTGAGCTTT GAAATCTC TTCTCTGAAA 180  
TTTGCATATC CACTCCAGTT GTGTACAAA AGATTTAAT CTCATGATG CAAATTCCTC 240  
TCTCCAGAA AAAATGACTA CAAAGCTC AAGGATATG CTTTGTGTGT CAAGGATTA 300  
CACTATGTT TTCTCTGTG TCAGATGCT ATTACAGAA GACTGTCA TCGAGAGAG 360  
TACTGACTA TCTTATGAC TTGTGATTTG ATCGAGCTT TAAAAAAA AAAA 414

60

(2) INFORMATION FOR SEQ ID NO: 18:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 469 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AATCAACAT GCATACAAA TGAATGCTT GTGAATGTT GACTGTGAC CCAATGCTT 60  
CAAGAACCTC GAGTGTAGAG TGTGTCTCT CCTGTGTGAC ATAGGTGAGA TAAATCAACC 120  
CTTCATATG TCAGAGCTTA GGAAGCTCTG GCAAGCTTG CCCCCTATT TGTTCGGCT 180  
GTTCGGCTTG CTTCGCCCGG GATGTAGCTT ACTGTCTCA GAGACGAAG GGGTTCCTT 240  
GCCAGAAC ATGAGGAGG CCAAGACTT TGGAGAAAA GAAAGCCCA AAGAAAAAC 300  
GAATTACTT AAGTCCAAA CTCAGAAC CTCGGCAC TGAAGAAAT GTTTTCGGC 360  
GATGTGCTT TGAAGAGAT AGATGAGAT TATCTCTGAG AAAAACTCT AAGAGCTTC 420  
ACTGTCTGT ATTCTCTCT ATACTGCTT ACCCCCAAT TAAATTCAG 469

(2) INFORMATION FOR SEQ ID NO: 19:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 550 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CCCCCCCC CCCCCACT TTCAGAGTC ACCCCCCAG ATTGTGGGTT GGGTTGACC 60  
TACTCAACC TGAAGCTGCC TGAAGAAC TCATCTCTT GCTCCATC CCCCCTACTG 120  
GTGCAAGAT GAGCTGTGA GTCGCGCTC TTGTGTGAG CTCGTCTG CCCCAGCCCC 180  
AAGCCCCAG CTCGTCTGA ATTGACATCA GTCTGTCTT GAACTGCTC CCCCACCT 240  
GGCATTAAT CAGAAACTT TTATGTTTC TAAAGCTTA GAGCTGCTG GACTCAAGG 300  
ACTGTGAGG GTAGCTGAG TTGCAGCTCA GTCTTGAGG GTCTGTAGG ATCTGACTG 360  
AAGACTTGC TACTCCCAA GCCAGACC ATCAGCAGG CTCGTGTGA GCAAGCTGCC 420  
TTGTGATGC TGAAGTAC CAAAGCTGAG CAGCTGTG GCTCATTTA AAGGATTTG 480  
ATGAGCCGAT GGGCTGTGA GCGAGCCAT TAAAGCATCT GGTCTGTTT TGGAAAAA 540  
AAAAAAA 550

60

## (2) INFORMATION FOR SEQ ID NO: 20:

5

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 741 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5 TCTTGAAGAG TGTACAGTAC AGGATATTA TATGTAAAGT TTAATTCAC AGGGTTTCGT 60  
15 TGGCTCTGCA TATATTATTA GCAAAAGAGA TTGTAAAGT GCCACAGTAT TCCAGATAC 120  
TTTTCAGTTG CGGCTTTCTT TCTGTTCTT TAATTGAAA CTTAGATACA TCCAGTAAAA 180  
20 ACTAGAGAAA TGACTTTTAC CTTGGGAG AGCAAGTTT TGTGTAAA CTTATTTCTT 240  
AGCATGCTT CAGCAAGTTG TCCAGACC TAGATTGTGA AGGACCACT GTTCTTCTGT 300  
TGTACAGGCT CCGTGAACCA TTGTTACAG GACCAATGTC ACATGCTTTC ATGGGCAATG 360  
25 NCCATGGGAG CATTGCGGTG ATATCTCTCT ACAGTATGCG CTCTTCTGG AGGCTGATAC 420  
ACAGGCTTC TCTTCCACAT GATCAATTCG AAACCTGCGC CAGCCCTTAC CATTCAATGT 480  
GGAGAGAAA CAAGACATCC CTGAAGAGA GTCCAAAGCTA CAGATACACA CCGTGTGCAT 540  
30 TCGGCTGTC ACTTCTGCC TCCCACTTCT GTATCTCTAG AGATCTGCG TCGATTTTTC 600  
CTTAACCTCA GCTGACTTCC CTGTGAATGT CTATGCTAG TTCAGGCTT CCAAGCATTC 660  
35 ATTGTACAG TGGTAACTCC CAATGAGGCT TCTGTATACA TTGTGTGTC TTATTTCTGC 720  
ATTAAGAAA ATGATTTTCC C 741

## (2) INFORMATION FOR SEQ ID NO: 21:

40

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 991 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

50 GGCACAGTCC TCCCTCTGGG AGTTTTTCT TTTTCAGGAG GCGAGAGGCG TTTCTGAGT 60  
55 ATGTGTGTTA GAGTGTGGG CAGAAATCT GGGACACAC CACACCAATT CTCTCTTAA 120  
TCCAGCTCAT TTGCTCTCTA TCCACGATAT GTTTCCAGTG TCTCTGGGT GTTTCCAGA 180  
GCACAGAAA ATGATTAAT CTCTGTGAG TTGTTTATTT GTTCTCACT TTGTTTACA 240  
60 CTGTATTTTC TGAATTATG GGTCTCTGAG AATTAAAGG GAATAGTAAA AATAAGTAAA 300

5 ACTCAGGTTG AAGGAATAT ACTAAATAA GATAAAGCTG ACCTGTAGAT ATAGCAGGTT 360  
ATAAAGCTTA GAGTGTGTTA AGTTGAGTGC AAATTTTCTT CTGATCTTTC TGATCCGAA 420  
CAAAAAGCA GTCTGTTTG TTATGTGATT GGAATGGAAC CCGAGAGAG AGCATGCTGT 480  
GTCTCTGAG GACAGGAAG CTGCGTGCA CCAAGTCTGA ACCACCACTT TCATGCTGAC 540  
10 ATAGATTATG TCGTGAACA TATTTACAC CGGCTTGCA GTAAACACTT GTATGTTGT 600  
CGAGTGAAGA CGGTCACTTT CCGCTAAAGC AGGGCTGTT GTGCHAGGGA AATGTCATC 660  
TCTGCTAAA ACACAGCTTC CATTGTATG TATGCTGCTT ACTCAAGAG TGTGCTCCA 720  
15 AACAGCTTT GGGAGTCTT CTTGTATTA TCGATGAAC CTGCAACATC TTGAGCACTG 780  
AGTTAACCAT AGTCTCTTAA ATAACTCTCC ACAGTTTTT CTTAGTTTAT CTCTACATCC 840  
20 AGGCTGTGCA GCGAGCTGTT CAAGTCATA TTTCTGGA ATATTTTCCA GTGTTTATTT 900  
GCACTTTAGC CCACTCTGTG TAGCTTATT TCTTCTAAAC TCACANTTAA TCTGATATAT 960  
25 ACTCAATTT AGGGGAGCTG TATTTGCTTT A 991

## (2) INFORMATION FOR SEQ ID NO: 22:

30

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 653 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

40 CCACGCTCC GGAATTCGCC TGAGATCTT GGGCTATCTT TGACAGGCGA TTCTTGCAG 60  
TTGATGCTTT CTACAGTGA ATATGATCAG TCCCAAGA TGGAGAGCTT GAGTTCTAC 120  
AGATTTGATG AAGATGGAGA AAACACACAG ATTGAGGATA CGGAACCCCT GTCTCCAGTT 180  
45 CTCATTTCTA AATTTGTTCC TCTGAAAAT GATAGTATCC TGATGATCC AGCACAGAT 240  
GTCAGATAC AACTGATCA GAATGATGAC AAACAAGS GAGATGATAC AGACACAGG 300  
GATCACATTA GTATTTTAC CACTGTTTCC AAGGCGAGAG AAGNAACGGT AGCAGAGAA 360  
50 GTTTGTATG ATCTACTTG TGAATCGGG AGTCAGGCGG TTGCTCAAC AGCTACTGA 420  
TCTGAGGCAC TTTCATGTT GTTAGATCAG GAGGAAGCTA TCGAATTTAA AGAACACCAT 480  
55 CCAGAGGAGG GGTCTTCAGG GTCTGAGTGG GAGAAATCC CTGACACACC TTGTGAAGT 540  
CAAGAGAGG AACTCAAGA AGAAATATG GAGATGTTT CATTGCACT TTCTCTGACT 600  
60 GAATCTAGT CCAAGGGTT GTGTCTTGG AGCATCCAA AAAAAAAAA AAA 653

## (2) INFORMATION FOR SEQ ID NO: 23:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1486 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

5 GCGAGCTGCA CAGCTGCA GCGAGCTG CTGCTCTG CTGCTCTGCA GCGGCGGAC  
15 CTGCGGACAG AACCTGCTG AACCCGAGA CCCCGCGGCG CATCGCGCG CATCTGCTG  
TTGCGCTGCG GACCTTGAGA GCGCGAAGC GCTTCTGCA GCGAGAGAG TCTTTCGCG  
20 GCTCGGCACT GATCTGCA GAGCTGCA TTGAGCTCAT GGTGCTCAT GCGAGAGAG  
TTGATTTGCG GCGGTCTGCG GCTGTCAATG CAGCTGCGCA CTTTTCAGAC ACAGTTTGC  
25 TGAATCCCAT GAGATTCCTG AACATCGCT ATGATGAGCT GGTGCGCGCG AACATGTCG  
GAGAGATCT TTGCTGTGCG ACTGAGATCA TTGATGCTCT CTGAGAGAGC CTGCTGTCG  
30 CTGTGTCGCG CCGGAGCTG GAGTTTACA AGCGAGAGA GCGAGTCAG AACCTTCATCA  
GATGTTCTTA CAAACATACT GCGAGAGAG AGTCTCATG TCGCGCGAG CATGAGCGCA  
35 AACATGTCAG GCTCATGCTG GAGATTCCTA AGATGCAAT GATCTCTCAT GAGCTGAGC  
AGATTCGAGC GCGCATCAGC GCGTTCGCGC AGATGCGGCT GAAACAGCG AACCTCGCG  
40 TTGATGCGCT GAGTGAAGCT GTTGAAGCTG CCGTGGCGCG GAGCGAGCTT CAGAGACCA  
GCGAGAGCTG CAGATGCTG GAGCGAGAG GAGAGAGAG GATCTTTC CAGATGCTGA  
45 GGTGTGAGAC ATGCTCTCTG CAGCTCTGCA CCGAGCTCTG AACAGAGAGC CTGAGTGA  
AGAGAGAGAG GCGGCGCGCG CTGAGAGTCG GTTGGCTGCG TGAATGCTCG TTGAGCGGAC  
GCTGCGTAAA GTGCGAGAGC GTTGGCGAGC CTGAGCGCGC AGGTGCGAGC ATGCTGAGCC  
50 CAGCTCTGCA TTACTCTATC AAAAAGACTC TCACTATCT GGTATGAGCG AGCTGAGCT  
CTGAGTGAAG AGTGAAGCG AGTGAAGCA GAGCTCTGCG GTTCTCTGCA AAAGAGTCAG  
TTGCGGCGCGG GATTCAGAGC TTGAGTTTACC GCTGCGCGAG CTGAGAGACT  
55 GGTATGAGCG CTGCGCGCGG GCTGAGTCG GCGCGCGCGG GTGCGAGAGC AAAGAGACA  
CGATTTTTCG AGAGCTGAG AGAGAGCTG GTGCGAGAGC AGAGTTTAA CTGAGAGAGC  
60 TGTGCTCTTA TCTTGTATAT AATGTTAAA GCGAGAGAAA AATTAAGAAA AAAAAAAA

## (2) INFORMATION FOR SEQ ID NO: 24:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2123 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

5 AACTCGAGCG GCGGCGCGGAC CGATGCGCC TTATGATAC GATANN  
1486  
10 CTTGCGGCTT CTGCTCTGCG GCGGAGGTTC CCGCTTCCCG TTGAGAGCTCC GAGAGAGCC  
15 CTGCGCGCTT CTGCTCTGCG ATCGATGTCG TCGCGCGCGC GCGCGCGCGG TCGCGCGCTC  
CTTTCAGCTT CTGCTCTGCG CCCGCTCCCG CCGAGCGAGC GAGAGTCGCG GCGATGAGGA  
20 TTGCGTCTCT GCTTCCGCGGA CCGGCTCCCG CCGAGCGAGC GAGAGTCGCG GCGATGAGGA  
AGAGAGAGC GCGAGAGCGC GAGAAATCA CCAATGATCT CTATGAGGCG CAGCATGTGA  
25 GAGAGATTA GAGAGGCGCA GTGATGGA CAAGGCTGCG GTTCCGAGGA TTGAGCTGCT  
GCTTACAGAG TCTCTCTGCT GCTGCGAATA ACAGCATTA GTTTGCTCT GAGAGAGTAC  
30 GCTTACAGAG TCTCTCTGCT GCTGCGAATA ACAGCATTA GTTTGCTCT GAGAGAGTAC  
GATATTCCTA TCGATGCTCT GTTAAATCTCT GATGCGGAGC AATGTCGTC ATGCGCTTAA  
CAGATTCCTC CAGATGCTCT CATGCGCGCG AACGATGCA AGTCTGAGC AGATTTTAC  
35 AGTCTCTGCG AGAGAGACTA AGCTTTGCT CCGAGTTTC TTTCGAGCT GCTCTGATAT  
TACCTTTTCT AATTTTATGA TTGAGTCTT TGTATTAAT TGTCTTCA TGAATCATTT  
40 TATCTTTTAA TTATGATTA AAGTTGCTCT CTATATTTAA AAAATTAAT ATTAATATAC  
ACAGAGCAT ATACATGCA AGTGAAGTC AAGAGCAAT CTTCAGAGCT AAAAGCATTA  
45 TTCTTATGA TTCTATATTA AATTTAGCTC TATGTATTT CTATGAGAGC TGAAGAGCGA  
GATCCGAGAT TTCTTAAAGC TTGATTTGAG CATGTCCTTA GTTACTGCTT GAAAGTGA  
50 TATATTTTC AGATGCTCT GAGAGCTCT ACTTTCGA TTGATTTTAT CAGTTGTAAA  
ATATGAGCA TGTGCTCTCT TGTGTCAAT TTGAGAGAAA AAAATTTT TTTCGAGCT  
CTAAAGAGAG TTGAGTCAG ATCAAGATTC TTCTGATAT TTACATCAT GCTGAGAGAA  
55 GCTTAAAGT TTATATGATA TCTTACGCT TTGAAGCTG ACTGAGAGAA CAAATGTCG  
AATTAAGCA AATTTATCT AGAGATGCT GAGAGCTAT TGTGTGAGAA AATGTATAT  
GTCTAAAGAG AGAGTCTTA ATGAGCTGCG TGAATATCT AACTGAGTT ACAAGTGAAG  
60 TTTTATGAGC ATGAGTAT TTGATGAAG GAGTCAAGC AATGAGCGA TGTGTCGAAA  
AGTATCTTT CCGCTCTTT GCTGTGCTCA TTGTGCTCTG CAGAAAGAT GCGCTGATG

5 CAGCAGGACC GCGAGCTGTA ATTAATAATA ATTGAGCTA TGAGACTAGC AAGGCACTAG 1380  
AACTGGAAAA GACCACAGAA AACAGAGAT CCAACCTTTT CATCTTACAG GTGACAAAC 1440  
TGTGATGATG CACATGATG TGTTTTGTAA CGTGTGAGCA CCGTACAAA ATGTAAATTT 1500  
GCCATTATTA GGAAGTCTGT GTGGCACTGA AGAGGACGCC AGGCGACTTG ACTCCGATGC 1560  
TGTGCGCCTG TCTACACGAG ACACACAG AGCTGGGTCA GATTCGCCCTC AGCTGCTTAA 1620  
CAAAAGTTCTT CCAACAGAAA GTGCTTACAA AGCTGCTTTC TGCGATACG AAGGTGCGAG 1680  
TTTTTGTGAC TCGACTGATT TTATTGCAAT TGAATAAAAA AAAAGCTAT TCCAAAGATT 1740  
15 TCAAGCTGTT CTGAGAGATC TTCTGATGCG TTACTTCTCT CAGAGGCAAT GTTTTACTTT 1800  
TATGCAATAT TCAATGTGTC CAGGANTAA AGTGAAGAAA CAGCAGCTTT TANTATATAG 1860  
GTCTCTCTGG AAGAGAGCTA AATTGAGAG AGAAAGCTGT GACAATTTTC ATATTCTCAT 1920  
TCTTAAAAAA CACTAATCTT ACTATACAAA AGTCTCTTTG AGAATAGCTT ACACACATG 1980  
GCCACAGCAG TTGTCTCTTA ATAGTATAGT GCGTATATCT ATGTAAATGCG TTACTACTA 2040  
25 CTGCGCTTAA AAAAAAACC CAGCATATTT ATTGAAAAA TGAACAGAA TTATATGTCG 2100  
TTAACCGATA TATTGTGCGA CTTAAAAAAT ACAATTAAAA CTGCTCTCTT GCTCTAGTAC 2160  
CATGCTTAGT GCAATGATT ATTCTATCT ACACATGATG CTGTGTCTTA TTTTATATA 2220  
30 TTTATCAGAG TGAATAAAAA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA 2280  
AAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA 2320

(2) INFORMATION FOR SEQ ID NO: 25:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 683 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

50 GGCACGAGCC TGTGTGTGTA TTTTCTCTGT GGTGCACTAC CTGACATGAG CAGCCAGCC 60  
TGAGTGCGTG AACAGCTTC CACAGGCTG CAGGTGTGTG TGTGTGTGAA AGAGAGAGGG 120  
GGCCGACAG CCGCTTTTGG AATGTTTGC CTGTCTGAC TGTGAGACAA CTTGGGAGTG 180  
ATTGTGTCT AATTTCACAC CTGCTCTGTT TTCTGTGACA TCTTGGAGG GAGCTATGTC 240  
55 CACACCATGC GGGGTGCTTA GAATATAAA AGTCCCGGGT CTGTCTCTCT CACTCTCGCT 300  
CTCATGGGGG AGGGAAAGA TGGCTTTGGT GCTTTTGTTC ACACAGCTCA TCGGTGCTGG 360  
60 GAAGGTGTGC ACAGTGAGCC TGTGTGAGG ACTGTGACCA GGTTCACAC TTGTCAACAT 420

5 CAGGCTTTTC TGTCTCTGAT AGGTTGAGC AAMAGTGAA AGAAGAGAA AGAGCTTTT 480  
CTCACAGCCA TTATATATA TAGTAGTGCG ATTCACTCTT CGTGCTCTG GCGACTTTC 540  
CGTGTGCTC AGTGACATGT AGATGACTGA CTGCAATATC TTGTCAACAT TCCCTGGAG 600  
CAGCTACTTA GGGGAACAA GATGTAGTGC TATTTGCGAT AACAGATGAG ATTTTCCACA 660  
10 CTAATAAAAA AAAAAAAAA AAA 683

(2) INFORMATION FOR SEQ ID NO: 26:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2036 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

25 CTGAGAAAG AAGGANTGCG GATCTGCTGC AAAAAACAT ATATCATAA AGACTCATGT 60  
TATTGAAA ACAGATTGTC AACACATCA CATTGCAATG AATCCTTTAA AAGCAAGAG 120  
ACCTTAAGT ATCTGCAAT CTGATTTTCT ATTTATCTCT TCAGTGAATA TAGAAACAT 180  
30 GGTATCTGA TTATTAGAGA TATATATTTG GATATGTTAC TTATTAATCT CTTATGCTG 240  
GTAACTGCA TAAAGTCTGT TATTAATAAC AACATAATTC TTTTATTTAA GAAGAAAGC 300  
35 TTATTTTCA TTGACAGTGT ATAGATTTAT CTACTTAGTT GTGTTTGTCT ATTAGTGTTT 360  
TAAATTTTTT TTTAAGTTGA GTGTTGATA AATTTTAGA CCGTGTGCCC ACCTGTGTTTT 420  
GAGTCTGTG TTGACTACAG GTATATAGCT CAAATTTAAA ATCTTAAAGC AAAAGATTT 480  
50 TATTTATAA AGAATCAAC NGTTGCAATG ATGAGGCTGT GAATCAGAT ATTTAGTAA 540  
AAGACAGCA GTGCTTTTTT TTGTATTTAC CCAATGACCC CCACCAATG CAACTGTTTTT 600  
65 AATATAGAA AATGATTAACA ATTTAAAT CTGAGAGTAA AATCTATTTC ACTAGATGCT 660  
TTTCCCGCTT TGTCTGATTT TAAGCAGTGT GTACTTGCCA TCTCTACATT GTCTTAGGA 720  
75 CAGTGTGTT CTACATATTT ATCATGTATG ATGTTTTTAT GTTCTTTTTT ATTCAATAG 780  
GCTTCTTACC AGAAGAGTA GGAAGAAACA CATTGACTGT GTACAGACA TGAACATTTG 840  
CTGCTGATAT GTTGTTTTTT CACTGCTTTT TGAGTTTTCA CTTTTTAAAC GAGAGCCAGC 900  
55 AAGCAATAA GATGTGCTG GGTGTGCTG TCCGGGCGGC TTTTTCGACC GAGCTCTCAA 960  
ATCTGTGTA TTGAGGTTTC CTTTTTGTGA CTCAGGATTT GAGCTACAGC TGGGGCCCCC 1020  
TGTCTCCAT TCGTTTGAG AGACACTGAG GGAACAGG GTTCTTTTG AGGTGCTCTT 1080

GGCTGCTTTT TACGGATGCG GAGCTTCTTC GCGATCTTTT GTTCTCTGCG AACCTCTGTA 1140  
GGTACGCGCG GTGCGAGGTT GTGAGTGTGA TTGCCCCAGA GCTTGGGCTK GGGGGCTGAG 1200  
CTGGGCTGAA TCGAAMAGCA TCGACACAGA AGCGGGGCGA GCGAGAGAAA AGCGAGCTCG 1260  
GCTGATTTGG TCCCTCGAGG ATGTCTGTGAG CATCTAGCTC GAGCTTGGCG CTGGGGAAAG 1320  
AGCTGAGACA AGCGGCTGAG GTGTGCTGT TACAGAGAGA AACCTGAGAA GAGTAATTTG 1380  
CAATGAGAGC TGTGATAGCA CTATGTGTA TTTTTTTTTT TTTTACAAAT CATGAGGAT 1440  
GTTTGCAAG TGAATTTTAT TTTTGTGTAA TTGCTTTATC TTTACTTTAA GGTGAAATGTG 1500  
TATTTCTCTG GAGAGATTAG GAGAGAAA GAGATGTGA TAAATGAAA CAGAAACT 1560  
CGTCCCTTAT TAAATATATA TCTGTATGTA TTTATGCCAT AATGTAACT GACTTTTAAA 1620  
AAGCTTCTTT TTTGTGCATG CGCTGTGAG GCACTGTAT TGTACATGCA TGCCTTTGCT 1680  
CGGTCTTTCG TGTATTAAGT TATGTAGAAA AGAAATATTT TTGCGCTAGT TCAATGTGCC 1740  
AAGCATGCA TATTTTATA ATTGTCTATA TATGTAGAAA GCAATTTGT TACATGTAAA 1800  
AGCTTACTG AATACAGAT AATCAATGT TTGAGAGAGC TGTCTTTGCG AAGGTAGAG 1860  
TTTTCAGAG TTTGTACAG TGAAGACCC TTGTGATTTA AAGTGTAC AGGTATTTA 1920  
TTATCATTT CCTCCATGT AACTAGAGAT CATGCTATA TTTCAATGCA ACGTAAAT 1980  
GAAATGAG GGAATGATT AATACAGCT TTTGTACAA AAAAAAATA AAAAA 2036

35 (2) INFORMATION FOR SEQ ID NO: 27:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 717 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GGCAGAGAT AACATAGGCA CAATATATCT GTATGTATAC TTCTAGATTT AATAGAGAT 60  
AACATAGAA TGAATTTTCC ATTGTAGAG AAAATATGTC GTTTCAGTGC CTTTATTTTG 120  
ATTCTGAG AGAGAGACT GCGACACAGA TTTCACCCCA GCGCTGATAT GAGATTAATC 180  
CTCAGAGCA GAGCCAGCA CAAAGAGCA ATGTAGAAA GTTACAAATG GAAAGTTTCC 240  
TGCAGCTTC GCGATAGCA CTCGAAACT GATGCCAGAA AGTCCAGAG TAAATTTGCT 300  
CAATAGCT CTACCAAGC AATTGAGCT CCCAAATTA ACTAGTGCAG TTGACTTAATC 360  
CTCTTACTT TTATCATTTA GGTAGAGCAT TGCAGAAAA CTCGTAGCTT TGCATTTA 420  
GCGCTGCTGT TCTCTGTGCT CGTGATTAG AGCGATGAGC AATATCTGGA AACATGAGT 480

AATACAGAT TCTTATCTTT CTCCGAGAC CTCTGATTT AAAAAATC AAGTTTCCA 540  
GGTATCTCA TACATCTTA AATTAGAAA CATTAGATA AAGTAAAGC ATTAGAGAGA 600  
GATTAATAG GATGCTGCG GTATCTTCTT AAGATTTCCG TGTATAGAG AAGTGCAGAG 660  
GTCACTGTAT CAGATTAAG AATATTTAT GTTAAACTA AAAAAATTA AAAAA 717

10 (2) INFORMATION FOR SEQ ID NO: 28:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 495 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GAATTCGCA CGAGAGCAT CTATATTTTA GTTGGAGT GCATTTAAA GATCTTTCTC 60  
TATCTCTTT TCTCCAGAA TTATCTTGA TTCTGCTGT CTGTGAGAT TTGATAGAG 120  
AATGAGCT TGTTTTAT AATTAATGA GAGCTGAT TTTTTCATTT CTGAGAGAAA 180  
AGCGATTTG CTCTAGAG CTGAAGGCA CCGCTGGTGT GCTGGAGCCC TCGTGGAGT 240  
TTCTGGGCGA TTGACCTTTA CAGATGAG AGCGCTTACA GAAAGACTG CAATTAATAA 300  
TTATTTTTTA AAAAGCTCC TCGAGGAAAT GCATTAAGG GGTAAAGCC GAGTATTTTG 360  
AGCTTGGAA GATGTAAAT AAGCTGAG AGAGAACTG AATCATGTAA GAGGTGGCA 420  
ATGATCATG GGTGCAGAC TATGAGTCC TCCGAGAGG TACAGACAT CAAGCCAGTA 480  
CAGAGAAATC CTAG 495

40 (2) INFORMATION FOR SEQ ID NO: 29:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 556 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AAGTAAAGT CAGATTTAT TACGGAGATG CAGCGAAA CATTAGTAG AATGCCCTTA 60  
GAGACTCTT AGAGAGCTG CTAGAAAGCG AGACAGACCC AAGCGCTTAA ATGAGATGG 120  
GCGACTGAG CTCTTCTGT GCTTATGCT AGCGGTGAG CAGATGTGTT CATCTGAG 180  
CATTAAGCT GAGATCATGT GAGCCAGAA CAGCGATTTT CTGGTTTACT TACAGAGAG 240

5 ANTAGAAGC AGGCAGATCT TTACAGACG TCTTACTON TTCCAAACA ATGGAAATCC 300  
CCACATGTCC ACUACACAGT KTGTTGTCTG CTTGTGCCAT GAGGACAGT GTGGCTGAGC 360  
GTACAGATC CCCACACTA AAGGAGCAG CAGATACAGS GCTGCACACT GTGTGATTC 420  
ACACATGTG CATTCTGAC AGGCAGATCC TGGATGGCAA AACGAGATC GGGCTGAGG 480  
10 GACTGTGAG AAGGGAGAG GGGCTGTCTG GATGTGGTT GATTGTAGCA GTAGCTCATG 540  
GAGATGTGAC CTCAAA 556

15

20 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 434 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 30:

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

30 CTAAATGTG ACTGTGGCTT TGTGAGACA GGGCCCAAT GTTAGTGTG AACACACAT 60  
GCACGAATG AGGAGCATG CAGAGTGCTG AATACTGTC CTGCACAGAT GTGTACATG 120  
ACTTTCTTTT CAGCTATTT CTGTGGCTG CCTTGAGCA TAGAGCTTTG TTGATATTTA 180  
CATTAACCA ANTGTATTA TTATGTCCA TTCTGACATG TTATTTAGCA AARGAAMAR 240  
35 GAGTAATCT ACATCAGAT CTTTAGTCA TGCTAAAGCA TTAAATGT CTTTTGGGGA 300  
ACATGTTTTG TATACATAA TGTTTAGATA GAATATTTA TAGAATCTC TATGTAGTA 360  
40 TTATCTCC TATGTATTT TATATCTAGA TGTGTCAATC TTTGTATGCA TATGAATCC 420  
TATGAATACT GAGA 434

45 (2) INFORMATION FOR SEQ ID NO: 31:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 715 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

55 CCAGGGTCC GATCTCAGC CTGACACT ATTCGAGCC ATACACACC TGTGTGTAGG 60  
AAGGTACTC CCAACTAAG CCAAGATGC AAGTTTGT TCAATGGGG TTAGACGCT 120  
60 ATGACTATCT CCAAAATCA CCTCTGGAT TTTTCCGAG ACTTGATGTT ATTGGTTTG 180

5 CTGGCTTAT TGGACTCTT TTGGCTAGAG GTTCAAAAT AAGAGACTA GTGTATCCG 240  
CTGGTTTCAT GGGATTAAGT GCTTCCCTCT ATTATCCACA ACAAGCCATC GTGTTTGGCC 300  
AGTCACTGG GAGAGATTA TATGACTGG GTTTACGAGG ATATATAGTC ATAGAAGATT 360  
TGTGGAGGA GACTTTCAA AAGCCAGGA ATGTGAGAA TTACCTGGA ACTAGTAGA 420  
10 AACTCCATG CTCTGCCATC TTATCACTT ATAGETAAC ATTGGAACTC CATAGATAA 480  
ATCAGTATTT CTACAGAAA ATGGCATAGA AGTCAGTATT GAATGTATTA AATTGGCTTT 540  
CTTCTCAGG AAAACTAGA CGAGCTCTT GTTATCTCT GTGAATCAT CCTACAGCA 600  
AACTAAGCTG GATCCCTTC ACTAGAGAT AATGTACAG CCTTAGACT CCTCATCTC 660  
ATGTTGCTAT TTATGTACT AATTAAGCC CAGTTAATA AAAAAAAAA AAAA 715

(2) INFORMATION FOR SEQ ID NO: 32:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 486 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

35 GAGCCATCC CGCGAAGG GAGCTTCTT CTACTTCTG CCACAGACC TGTCCACACA 60  
CAGTTCTGC CCTCTCTG CTGGAGGCC ACTTCTCTCC CGAGTGTGG ATTCAGCCCC 120  
CAGCTCACC TCAAACTGG CCCCCTCTCT CTCTCTCTT GCGCTCTCT GCTCCCTGA 180  
GGCTTTCTG TCTCCCTTC TTGAAAAGCA ATGCACGCTT CTTGGGATCT TCTGCCACT 240  
40 CAGCTACCA TGCCCTTTGC TCTGTGAGC TGAGCTCTC AAGGGAAATG TCTATCTCG 300  
GTGTCTGCT TCCCTCCCTC AACTCTCTCA CCTGTCTCA AGCTGGCATC TCGCCCTCA 360  
45 CTCACAGAA CGATCCCTC ACCACTGCC TTATACAGGA GAGAGCACA ACNTGAGAA 420  
AAGAACTAT AGGGCTACA AAGATGTCTA CCTCTGATCC CGAAGGCAAA AAGNATCTTT 480  
GGGAC 486

(2) INFORMATION FOR SEQ ID NO: 33:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 725 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

5 GTTCCCTGCG TAAATATGAG GTTATTCOCA GAAACACAGT GTCAATCTTT AAAATTAAGC 60  
TTTCCCTGTT AAACCTTTTC AAGAGACAG ACCACCTTCA AGATTCACCC TACGGTTGAT 120  
ATGTCTTAA TTCAATTTAT AAAATTTAT CTGTCTTCA TTTAAGCT TTGGCTAATAT 180  
10 AGTCAGAAAT GTCCATAATA ACAAACTAT TTGTATTAA TTAGAGAG AGTAAGGGA 240  
AGAAAATGA AAATCAGATC TTATATGAG CTCACAGAT ATTAGGCTT AAAGGCTTT 300  
TCATGTTTA TGAATATTC TACTACTAT TTTTATATAT TCGTGTTTT GATGAACGA 360  
15 TCTCTGGGA AATTGTGAG TTACATAGC ATTTCACGT GATTCCTTC AACCTGAGAT 420  
CAATCTATA ACCAATGAC AACCTCTC TTGGTTTAC TGTCTGTGA AATGACAGT 480  
20 CAAGTTGCC AGAAGTCCTG TGTATTGAT GATGACAGT GCTTTCTTC GTGGAGAG 540  
TTCTGACCC TCTTAATTT GTGTATATG CTCACAGTA TCTAACTTC CATCTGATC 600  
25 TGTATATCT ATCCATACG TTAATGTAT TATGATATAT GTTATATATC TTGCTTGAG 660  
GTTCATCTT TTCATTTGA TGAATAAAA GTTTTCTCT GCTATAAAA AAAAAAAAA 720  
AAAAA 725

(2) INFORMATION FOR SEQ ID NO: 34:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 437 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

5 CACACAGAT GCTGCCCTCA GACCTGTCA TCTGTACGA CATTGAAAG CTCCTGCTCC 60  
45 TCCAGATATC TGAAGATTC AAGCATCTC TCGAATGTT CCGAATAT TCCAGATATC 120  
CTCTCTCTT TATTGACTT TGTGAAATC AAATTGAGG ATCTCAGTC CCACTGAC 180  
50 CCAAGTGGG AAATATTC AGTTCATCC CCAAGAAAC AACCTGATC ACATGACTC 240  
AGAAATCTTA TAACTATGAT GATCTTTTC CATCTGTACA TTGTCTGTGA CATGCACTC 300  
ATCACCTGAC GTGCCAGAT CCTGCARAG CAACACCTG TGAATATTC AGGTATTTT 360  
55 CTACATCTAC AGCTTATGAT TACCTGATA TCACTTACA TTCTACATAT AAAAGAAAA 420  
AAAAAAAAA AACTGTA 437

(2) INFORMATION FOR SEQ ID NO: 35:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 943 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

5 GGCAGAGCT GGAACAGGA CTAAATCCA GAAATGAC ATTGTATAC ACATTAAGC 60  
10 AGAATAGCT ACCTGTGGA GATTATATAT ATATGGTTC ACATGATCA TGTATATAT 120  
15 GATGACAGG ATGCTTATAG TGTACATTT CCTTATCTT GTTAGATTA CCTGCTTAA 180  
20 TCTTGTCCA TTACATGCT AATTCCTCT CAGTGTAT TTCTAGTGA CAGATCTTA 240  
25 ACATTTCTTA CACCTGGGA GAAAGGAGG AATGTGTT TGGGGTGGT AACTAATTT 300  
TTGATGTAA TTGCTATAGA TGAATATGA AAATATGGA CAGAAATGT TATTAAGAAA 360  
30 TTAACATGAC TTATAGTGA AAAATATG CCAATAGT ATGTCTTCA AAAGGAAA 420  
25 TTAACTTTC TTTTATACC ATTGACTG CTCCTTAAT AGCTTACAG ACATTCAGT 480  
TTAACTGAC TTATAGTGA AAAATATG CCAATAGT ATGTCTTCA AAAGGAAA 480  
30 TGTCTTGGG TGGCTTAAAG CTAAATTTG GTCATTTAC ATGAGAGATG TTGTATGAT 540  
TGTCTTGGG TGGCTTAAAG CTAAATTTG GTCATTTAC ATGAGAGATG TTGTATGAT 540  
30 TGCATATAT ACGACATAT TTATCATAG TGTATTTT TGTATAGAT TTTTATACC 600  
ACTATATGT TGAATCTTT TTGTCTATN AGGTGAAAT ATGACATGC TAAATTCAAA 660  
35 CAAATTTCC ATTTGAGAA CTAAAGGAG CAATTTAT ATGAAGAAA TTATGAAAT 720  
ATGATATTC AACCTTGA TTAACTTGA GATTAATAG TAAATATG TGGAGAGAT 780  
TTGTGTTTC TGAATTTT GTGATTTAG GAATATAT TACCTTTAT TTGTGATTA 840  
40 AGTATATGT ATGTGTATA TTATATATA AACGTATCC AAGTAAAAA AAAAAAAAA 900  
NNAAAAAAAA AAAAAAAAA AAAAAAAAA AAA 943

(2) INFORMATION FOR SEQ ID NO: 36:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 604 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

55 GGCAGAGA ATCTTATAC TGTATGACT CCAAGCATG GATGAGCTTT CCAATGATAT 60  
60 GATCTTATG TCTGACCTGC AGTGTGTTGG TTTCATAT TCTTCTATAT TTTTCTTCA 120

TTGATCTTG GGAATTTCT TTGTTGCTC CTGTGTTTC CAGCTTTAA TAAACACAG 180  
CCCAACAAA AACATAGCA TTCTGACAA TAGGGGGCC ACATTGGACC CAGTATGTCA 240  
5 CTTPAATGCA CTTCAAGAA AATCTGAAT GCGAAAAATG ACATAGGAA TGTATCTCC 300  
ACACATTTTA TCCCATATA TGGTGTGTT TCTTAATTTT GTTCTGTG CCGAAATGTG 360  
GCTTTCMAAT TAAATGACC TTGTTCTCTT TGAACCTTTT TGTGTTGACT TGTATATTA 420  
10 AGGTTTGGCA AAGATTGATA ATTCTGAG AGGTTTGGCA CCAGGAGATA CAAGAGATC 480  
TCAGTAGTAA TCTGTTTCAT GTGCTTTTAC AGCCAGCTAC ATTAAAGGAT GTATATGTTA 540  
15 CAGAAATTAAT ATGTCTGTGT ATGTGTTCTT ACTCAATTA GTACATGCTT CCAAGAAAA 600  
AAAA 604

20

## (2) INFORMATION FOR SEQ ID NO: 37:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 349 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GTGATGCCC GGGAGCCCC AGGCGCTGCC CCTGAGAGG ATATCTTTTA CGCTGCTTT 60  
GTCCACACC TAAACCCCCA GCTGCTGAG CAGTGGGCAC ATGGCAAGGG CTCACATGG 120  
35 GGCACATAGA GCATTTGGGG GACTGGGAGT GCTCACCTTT GACTTCCTGC AGGTGCGGG 180  
AAGAACAGAT CACTGATACC AAGCTTACA TATCTTTGAT CTTCACTGTT CTGATCTCTC 240  
40 CCTGCTGGG TCTCACACAG TATATGCCAC CACTTCTCTG TCTAAATTTCA GAATAGAGT 300  
CACATCAGGA GAGCACTGTC CCGAGANNA TCGAAGCGG TTGGAGCA 349

45

## (2) INFORMATION FOR SEQ ID NO: 38:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 672 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GTAGTGGTGG CGGTGGCGG GATGGCGAG ATCTGCGGCT TTGAGTGGCT AAGAGCGACC 60  
TGGTAGCCGG TCGTATGTCG TTGTTGATTT GTATTCGTTT CTACAGAGCT GATGTGTCCA 120

60

GGAAGCGAG CAGCGTCCAC AGCGAAGTAA ATAGTAATCC CCGACCAAGT TTCTCTCGGC 180  
TTTATCATGT CACCCACTGT GGTATATGCG TTGTGTTCTG CCACTTTTCG CGTGAACAA 240  
5 TTCAAGATA ATCAGATGCC GCGTGGCGCA ATATTCACGA TAAAGCCCTGG CAGTGTGGG 300  
GCTGATGTT CAGTGGCTGC GSCACCGTTT YTGCGGTATG TTGCACACCA GEMTCTTTAA 360  
ACAGTTTTCG SAGCGGTTT AGGTCAMAG GTTCAATGCC GGTGCGTAGC TGTCTCTTAG 420  
10 GTTCACCGCG AGCATAGCA TTAAACATCT CATCAATTTG CTTCTGGCTG GCGCTATCMA 480  
TACTTTCCAG CATATGTTTA CGCTGGCGCA AACGGTTAG CGTTTGGCCC ARCMGATCAT 540  
15 AGCAATGCG CTTAATGAGA TAATCAATA CACACACAG TAGGCGTTTCA GACACCGTTT 600  
CCATATGCT GCGTGGAGTG GTAAACACCA CGTGGCGGG ATATGCGCC TGCACCAATT 660  
CATGCAGTAA AT 672

20

## (2) INFORMATION FOR SEQ ID NO: 39:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1908 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

AGATTGATA TTTTAGAAA CAGTAATTTT ACTTTTAGG AATTTGCTA GCTCTTTGAC 60  
120 TTAGAGCTG TAGGAGCTC AACATTTCTT TTAGAGAAC GTTCTTTT TTGATTTGTA  
180 CAGTATATA AACATGCTT TTGTTGATTT GTATAGGTTT AAAAAAGAA TAACTGTATG  
240 CAGTTTGA AAGGAATGT GCTTTAGCA TCACTCATAA GATGCCATG TACTTTAGG  
300 CATTTATTT TCGTTTGA AATGACATCA GCTCTTCTCT TCTGACTGT AACAATAGC  
360 CCGAAGGAT GAGATTTT TTCAATGGT TTTTATGTT GTTATGTTT GGTTTGTAC  
420 CCGAGCGAG TGTGTTGCG GACACTGAC TCTGTCTCT AATGAGAAC AAGTTAGAAA  
480 TCTGCGATA ACCTAATAA ATTTAGAAAT GAATTAATAA TGTGAATG GGTAAAGTG  
540 ATGATGATA ATAGCATGC AAAAAAAG CTCCTTCAT CAGACTGCG TACTGTTTC  
600 TTCTGTAGC ATTTGTTTG GAGAGCTC TTGTTTCTT CTTTGGGG TATGTTCTG  
660 TTCTTANTA TGTGTTAAC ATTAATGAGA TATAATCAC ATACCTTACA ATCACTTAT  
720 TTTAGGATA CAATTTAGT GTTTTTAGT TATTCAGAA GTTGTATAC CGTGACACA  
780 GTCAATTTA GACATTTG TTAACCCAAA AAAAAOCT GTACCTTGA CGAGTCAGCT  
840 CTCATTTCT CCGATGCCC ACCCATCC CAGGCGCGG GAACCACTAA TCTATTTCTC



TCCTCTTACG TTTCTTATT CTCTGATT CATATTAAG GATCTTACG ATTTCGGTC 900  
TTTTCGGACT GGCCTCCGCA ATATGATTCT CTATATCGAG TCGAAAACT CTTCGACAT 960  
5 TCGAACTCT TATCTCTGAG AAGCGAGTC GTTGTAAAA TCAATGATT TCGACGAGA  
GGCCGAAATC CTTACAGGAT TTTTCTCGCT GAACTTAATG CTAGATGATT TTTCTCTGTA 1080  
10 AACAGGATT TATGATTAAT TACTACTAT AATATCTGAG CTATCTCTCA GAATCGAATG  
TGAATCTTT GGCCTCTGAG TCGGAGAGG TCTCTTAAA AATCGAGCTA AAAATTCAGGC 1200  
GAAAAATCAG GCGAAATGAC TTGCGAATA ATTGACAAA TGGTTTTCAC GTTGTCTCAT 1260  
15 CTTCCTACG AGCTTGTATA CCGTAGGCA GGTGAGCTCC GCAATTTCT TTTTCAATT  
ACTCGAGTA GTTCTCTGTC TCTTTTCA GTATGTACCA TACGACTTAA AGGTGATTTC 1380  
20 GATCGGTTGT AACCTGCTA AATATCTTA GTACGAAAT TATCTACAG TACTGTGACA  
GATGAAATTA TTGCTAGAGG TATGTCAAA ATATGATGTC AGCTATTTAA GCGTTTGCCT 1500  
GACTGATTTC AGTATATTTC AAGATGTGT ACCAATCGAG GCTCTTTAAA ATACGGAAGC 1560  
25 ACTGACTTAA TAAAGAGGGA ACCAGCGAAA TACTGTGAG CCGAGAAATA TCGATATCAA  
TAAATGAGAG GTATTTATAT TCTGTAACTC CCGAATGATT GTTTCGTAAG CATGTGAGCT 1680  
30 TCTGAGTGC TTGACAGCAT CTTCCTGATT GTATGTGAGC TTTTACATG ATGTATTGAA 1740  
TAAATGTGTT TGTGTGAGC ATCAATGCTT GTATACGCA ACTAAACAG TGTTTTTCGG 1800  
AATATTTTC AATCTTAAA TAACTTGC CTATCGAATA AATAAATGAT TGTCTACACC 1860  
35 TGTAAAAA AAAAAAATT AAAAAACTG GAGCGGGGCG CCGGTACG 1908

40 (2) INFORMATION FOR SEQ ID NO: 40:

(1) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 458 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CTTCAAAAAA AAAAAAGAAA GGAAGAGGT CTCTACAGAA GCGCGGATT CTTCATGCGA 60  
AGGATTAACA TCGAATAATG TTCAATTTCK GCAATTAAGTT TCGATCTCT TCCCGATCCA 120  
55 GCGATTAAGA GAAACGAAAG ACAATGATG TATCTCTGT GTCTCAGCT TTGGCAGCTT 180  
TGTGATGTT GCTAAGGAGC AGTGACCTTG CTAAAAAGAC TGAATTAATC ACCAGTGA 240  
60 TACCTAAGCT GGGGAGGAAA TGAATATTTC CTTCGTGAT CTCCCAAT CCAATGTGTT 300

GACGAGGCCC TCCGAGAGC TCTCTAGTTC CTTCAGATG GAACTGTGTC TACTTGCCCC 360  
GGAAGCGAAT AGTATGTGTC CTATCTGAC CTTCGATGAG CAATGCGGCT GCGCTGTGAA 420  
5 CATTAAGCTT CATTTTTTTA AATAACTTGA AATTGAAA 458

10 (2) INFORMATION FOR SEQ ID NO: 41:

(1) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 1153 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

20 GGCACAGAGC CTCGAGCCA GGTGTCTGAG AGCTTGGGAG GAAAGTGTG GCAATCGAGA 60  
GGCTGATCTT GAACTGTGAT TGGGGAGAGT GGAAGCTGTT TTAACGATGT GCGCCCTCTC 120  
CTGTGCGGCG GTGGGCAATCC CCGGGGGCAG TGGAAACCGG GGCCTCTGCG AGCTTCCGAG 180  
25 TCGAGCAGC CTGGGCGGCG GCGCGGCGCC GAGAGACCGG AAGATGCTCT TCCCTCTGAG 240  
TAACTGTAC TGTGAGACTG GTCTTTGTG GCTCAAGCCC GTGCGAGAT TGAAGCTTAC 300  
30 CTGCGAGAT GGCACAGAGG GGTGAGAGG AGAGAGGAGG GCAATGAGCG AGCTTGAAGA 360  
ATCCGTGCTC CAATCTTAC ACTGAGATG GCACTTGGCA ACTTGTGTG GCAATGAGCTG 420  
GGGCAAGAT CATTGAGATT TTAACGAGAG AAGAAATCTC GCTGACCAAT CTCAGACAG 480  
35 TGCATCCCAT GGAATGTGTC GAGGTGAGG GTGCAATGTG TAACTTGCAG CAGCGAGCAG 540  
GAGCGACAGA TCTCCGAGGA GAAGTGTGAT GAAATACTT GCGAGAGAT CATTCACATC 600  
40 GTGAGATGTA TGAATCGGCA TGAATACTG CCGAAGATGC CGACAGACTC GAACTGTGAT 660  
AAGCTGTGTC ACACAGGCTT GCGGAGATG CAGCCGACCC TGTACAGAT CTCTCCGAG 720  
ATCACTAGTG CCGTGGGAC CTCAATGAC ACCAATGAG GAGAGCTCAT CAAGAAGAC 780  
45 CTGCGCTGTC AGCTGTGTC GATCTGTGAG AGCTCTTGA TGGATCCGAG AGCTTGCATT 840  
TTGGAGATTG CACTTTTGGG CATTGAGGCT CTGAGACTG CTCTGAGTA TTGGTGAAGAC 900  
50 TTGAGAGGAG CAGCCCCGCG TGGCTTCTTG GTTTTGTGAT TCGCAGACTC AGCTTATCTT 960  
TTTAATCTTT GCTTACGATT CAGTCTGAG TCTACTGCT CTCAATAGCC CTGATGAGCT 1020  
CCTCTCTCT TCTCAGCTAT ACAGAGAGC CAGCAGTGG AATGGAAATA AAGTGTAGAA 1080  
55 CATTGATTG GCTTGAATAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA 1140  
AAAAAAAA AAA 1153

## (2) INFORMATION FOR SEQ ID NO: 42:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1983 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

5 GGCACGAGAG GGGCCGAGCC GACAGATGT TCTTCTCTCC TCTTCTCTCC GCGGGGGGAG 60  
15 TAGTCTCTCG AGCTCTCTCC GTGAGAGCTT TGGGAGCCG GAGTCTCTCC ACCCGAGACA 120  
TCACGAGGGC CCTCTCTTTA GGATCTATT CCAAGGAAA AGAAGATGAT GTGCCACAGT 180  
TCACAGTCC AGGAGAGAT TTTGATAAAT TGTTAGCTGG AAAGCTGAGA GAGACTTTGA 240  
ACATATCTCG ACCACTCTTG AGGCGAGGGA AGACTCGAAC CTTTATGTGT CTGCCTCAGG 300  
ACTTCCCGAG CTTGTCTCTA GTTGGCTCTG GCAAAAAGGC AGCTGGATC GAGCGACAGG 360  
25 AAAACTGGCA TGAAGCGAAA GAAACATCA GAGCTCTCTG TCGAGCGGG TCGAGCGAGA 420  
TTCAAGACCT GGAGCTCTCG TCTGTGARG TGGATCCCTG TGGAGACCT CAGCTCTCTG 480  
CGGAGGAGC GTTGTCTCTG CTTATGAT AGCATGACCT AAACAAAA AGAAGATGG 540  
30 CTGTCTCTCG AAAGCTCTAT GGAGTGTGGG ATCAGGAGGC CTGGCAGAAA GGAGCTCTGT 600  
TTGCTCTCTG GCAGACCTTG GCAACCCAT TGAATGAGAC GCGAGCCAT GAGATGAGC 660  
35 CAGCAGATT TCGCGAATT ATTGAGAGA ATCTCAAAAG TCTGTAGTGT AAACCGAGG 720  
TCCATATCAG ACCAAGCTCT TGGATTGAGG AACAGGCAAT GGATCATTC CTCAGTGTGG 780  
CCAAAGGATC TGAAGAGCCC CAGTCTCTCT TGGAAATTTCA CTACAAAGGC AGCCCCCATG 840  
40 CAAAGCAACC ACCCTCTCTG TTTGTGTGGG AAGGANTTAC CTTTGAAGCT GTGTGTATCT 900  
CCATCAAGGC TTCTGCATAT ATGCACTCTA TGAAGGCTGA CATTGGAGGA GCTGCACATA 960  
45 TATCTCTCAG CATCTGTCTCT GCTGCNAAG TTAATTTGCC CATTAATAT ATAGGCTCTGG 1020  
CCCCCTCTTG TGAATATATG CCGACCGGCA AGCCGACAJA GCGCGGGGAT GTTCTTGAAG 1080  
CCAAAACCG GAGACCATC CAGTTTGAATA ACATGATGC TGAAGGAGG CTCATCTGG 1140  
50 CTGATGCGCT CTGTATAGCA CACAGCTTTA ACCCGAAGT CATCTCAAT GCGCCCACT 1200  
TAACAGGTGC CATGATGATA GCTTTGGGAT CAGGTGCCAC CAGGTCTCTT ACCATTCAT 1260  
55 CCTGCTCTTG GAAACATC TTGAGGCCA GCAATGNAAC AGGGAGCCT GTCTGGAGGA 1320  
TGCCTCTCTT CGAATATAT ACAGACAGG TTGTAGATTT CCAAGCTTCT CATTTTACA 1380  
60 ACATTTGAAA ATACAGATCT CGAGAGGAT GTACAGCTGC AGCATCTCTG AAGATTTGG 1440

TAATCATTC TAGTGGGCA CATTAGACA TAGAGGGGT GATGACGAC AAAGATGAG 1500  
TTCCCTATCT ACCGAAGGC ATGACTGGGA GCGCCACAG GACTCTCATT GAGTTCTTAC 1560  
5 TTGCTTTGAG TCAAGACAT CTTAGTTCA GATACTCAA AATGTCTTCA CTCTGTCTTA 1620  
AATGGACAG TTCAACTTAA AGGTTTTTG AATAAATGA TGAATATCTT TTAACCGAGA 1680  
CAAGGATGG TATTTAAAA TGTAGACAC AATGAATTT GTATGCTTG ATTTTITTTT 1740  
10 CATTTACAC AAGATTTAT AAAGTAAAG TTAATATCTT ACTTGATAG GATTTTAAAG 1800  
ATATCTATA AATGATTAA ATTTTAAAG CTCTCTATC ACTTTTACA GTATATGTTT 1860  
15 TTCAATGGA AGCAAAATG TAACTCAGT TTGTGATCT AGGACATGA GCAACTGAA 1920  
AATTAATG CACTGTGAG AAACATAAA TCCAATGTT TGTCCAAAA AAAAAAATA 1980  
AAA 1983

## (2) INFORMATION FOR SEQ ID NO: 43:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1406 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATGATGATCA CTTTGAAGAC GATTTTATC CTCTCTCTCC AGCTAAGGC CTTCAGGTTA 60  
120 ATAGTTGAAA AAGACTCTAT AGATATGAC ATTTCTTCA GAGAGAGAGA AGATCAGTCT  
TTAAGGCTTA ATGCTTAAGC NCTGTGCTT AACTGACCT GCGATAACTA CTTTAAAGAA 180  
240 ATAAAAAAT CAGTCAAT ATTCTCAAC TGAAGTTTA GTGGCAGCAG TTCTATTTGC  
CTTCACTTA TCAGCTACT ATTGTAGAAA GTGTAGACA TACTGACTCA ATTTCTTAAGT 300  
360 CTGATTTGTC CAATTTTITA TGTACTTTT TAATAGCT TCTTACGTGC AATCTGAT  
TAGAGTTAAA GCGCTGTGT AAATAAAGG CTCAGCAAA ATTGTACAGT GATAGCAACT 420  
480 TTCCACACAG GAGGTGAAA ACAGTAATGT GGTACACAG TTTTITTAAC TGTAGAGCA  
540 TCAGCTGCT CTTTATNTA TGACTAACA ATATTTTAA ACAATCATA GTACGACAT  
600 ATTAAGGCTT TCTATATGC TAAATACCC AGCAATGATC TTGTGCTTTT TGAATTTAT 660  
720 GGTAGATGTT TCCCTCTG AGTTTGTCA GTTTCAGCT GTTTCCTGCG CCAGGTGTAC  
780 TGTTTGTGCG CTTTGTAT ATCCCAACC ATTGTTGGG AGTCAGATG GTTCTTTAAA  
840 AAAAAAATA AAACGACAT ACCTGACAG TCACTTTTC GTTCATTATA TGTACCGAG  
900 GTACAGTGT GTGGGATGAG GTTCGATACA GACGATTTTA TTGCTGTGCA TGTAAATTTA

AAACCTGTA TTAACTCTT TGCATCCCTT TTAGATATAA TTGCTCTTGG CAGACAAAT  
900  
TGTGCTTAT TTTTTCAAA ATTGCTGTG AACAGCTGA TGCACACAG CACCTTTAT  
960  
CTAATGAACT AACGCTATCT TAAATGCTT CTTCAGAGTT TGTGTGTGAC TTGTAAAAAG  
1020  
CTACAGGAA TATTAATAAA ATCTATCAC TTAACTTAT AATGATTTAT GAAATATAAA  
1080  
CATGATCAC AACTTTGTT CTGTGTGAC CTATATAAAA AGTTGTCTT TGCATTTCA  
1140  
TGTAAAGAC TGAACAAAT GTTATGTGT TAAATTTTG TGTGTGTGA GAAATTTTG  
1200  
TCAATAGAA TTAAAGAGAC TTACAGGAA GATTTTAAAG TTAGAAATAT TCCATGCGA  
1260  
TAAATATGA AATTAATAT AATAGATTTT AACCTGCAAT CAGTGGAAAT CTGTGCTAAG  
1320  
TATGTATGTA GTTATTTATG AACAAGAG ATTTTTTTGG CTATTTACAG TACCAAAAG  
1380  
GACCAATTA ATGGTTTGA GACTTG  
1406

25 (2) INFORMATION FOR SEQ ID NO: 44:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1391 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGCTGAAG GCGGACCGC AGTCCGAGC AGTCTGCTT CTTCTGCGG GCGCTGCGGC  
60  
CCCGGCGCTT GCGAATGACA GTGAGCTGGA CATTCTGTG GGAACACGA CCGTTATGGA  
120  
CGAGAGCTG TATGCGCTCT GCTCGAATG TTACTGTGTG AACAGCGGG TGGCGCTGCG  
180  
GGTGGCTGCG GGAATCTGCG AACAGACTG CCGACAGCGA GCGGTGCTGC AAGACGACAC  
240  
CATGAGCAAT TACGCACTT TCCATGCTT CAGAGCGCTG CTGATGCGC GCGCCAACT  
300  
ACTGACACAG CTGATCTTCC AAGATTTGCG CTTCCCGGAG GCACTACTCA TCGAGAGTTA  
360  
CTATGCTTTT GATGAGGCTT TTGTGTGGA GTTGTGCGG AAGAACTGT CCAAGCGCAC  
420  
CAGAGAGAC CTGAGTACA TCGACACGA AACAGCAAT ACCCTCAAGA GTGTGCGGAG  
480  
AAGTTTGAAC AACTTTAAAC GGTCTTTCA GGTGTGTAGG GAAATCGGG GCTTCCCTGT  
540  
GAGCAATAT CAGACAACT TGTCTCTTC TGAACGTTTG GCGAGGAACT ATTCAGCAAT  
600  
CGTCTCTTT GCTAACAC GCTTTGAGAC AAGGAAAGAA AAATGCAAT ATTCAGACTT  
660  
CGGTAACTTT GCTTTCTGCG CTGAGCTAT GATTCAAAC TGAACCTTG GACCGCTGCA  
720  
CTCAGAGAT GATGACATG AACATGAACTT AAGACGAAT TTCTCAGGA CTTCAGAGAG  
780

CTCAGAGTGC TATGAGCTGA CAGAGAACTT CTGAGCTGCG AAGAGAACT GGTGTGAACT  
840  
GCTTCCCGGG AAAGCTGGGC GTCTTCTG AAGTGGAAAC CAATTTGAG AACCTGTCCC  
900  
GGGGGCTGTGT GAAATGCGCG CAGAGCTGAC CCAAGATTAAG GATGTGAGAG AACCTTTGT  
960  
GAACTGTGAG GAAATTTTG TGAACCTG CCGCTTCCAG CAGTGGCAC TGCAGCAAT  
1020  
GCGGTCTTC CTGAATCAAT ATTGAGCTTC TTGTCAATCC CTGATGCTT TCCACACGA  
1080  
GCGCTCTGCG GACGCGTACA TGGGCAACCTT CCGCGGCTGCG CTGTGTGCGG TGTATCTAGA  
1140  
CTGAGTGGCC TCCCAAGTTC CGCCCAAGCT GACATTAAG TTGCTGTAGG TTGTGAACT  
1200  
GCTCTGCTCT CCGGGAGAGA AATGGGGGGC GTGCAGATGT GCTGTGTCT GTCTGTGAGC  
1260  
AAGTGTGTG CATTGACAG GATGATGTG TACATGCTCT CATTGGGAG TGAAGACTAT  
1320  
CTCAGGAAAT GGTGTGTGAG CTGAGCCCAAT CAGACGAGAG AGTGTCTCA CAAAAAATA  
1380  
AAAAAAAAA A  
1391

25 (2) INFORMATION FOR SEQ ID NO: 45:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1569 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GGACAGATG GAAATGCTG CCGCTGTGCG GGGAAATGCG CAGAGGATTC TTCTGCGCCA  
60  
GGGGGGCGGG CTGCGCTAGCC TGCAGACTGT CCGCTATGCG TGCAGAGCTT TTACCGGCGA  
120  
CGTCTGTGAG ATGCACTTTC ACGGCGAAGA GCTGAATGCT GTGACTGAAT AATATCCGCC  
180  
GAAACAGACC ATGCACCAAT CATGCTGCGC AATCTCTGCC AACCCCCAC AAGAGAGAAAT  
240  
AAGCTGTATC AAGCTTCTGC GCGGGAGAAAT AAGACAAATT TTCCAGACA ACCGAATGAT  
300  
AACCTGTGCG CAGATGTGAG CTGTGATGCG AAGAGACAG CTTCCTTAATG CAGACAGAGC  
360  
TCCGGAAGA CAGATCTGAG ATGAGGCTCT TTCCACACA GTTCTTGAAG GCGCTTCTG  
420  
GAGATTTCA AATTAACAAA TTGTGTGCGC CTTTTTTGG GCGACAAAT GTGTGCTGTC  
480  
AATGAAGAG CCAAGGTACA GGAATGTGTA CCAATTTTAA GGAATGTGCG CATTCTGCGC  
540  
GCTGTAGAT GGTGTGATG ATGACCAAT CTTCAGACAG CAGGCTTTTA TCAATACTTC  
600  
CAAGCTCCCC AAGCTGTGCC TTGTGAGAGG GAGATGTGTA GAGAGCTTCA CCTGCGTAC  
660  
AAGCCAGACC CAGTCCGCGC TTCAAGACACA GCGGCTGAG CTGACAGCC TTGTGAGACA  
720  
GTACATTTGA GAGCAGAGCG AAGAAAGATT CTGTATGTC GACCAATGCG AAGCCAGATC  
780

CTGACACTGT TCGGACCTCG TAGCAGCCT GTTATGCCAG CCTTCGCCAT AATATCACTC 840  
TGGGTATTGT GCTGTGCTGT COTCAATGGG ACATGTGGAA GAACTTGGGG TCGGGGAGTG 900  
TGTATTGCAC TTGGTATTCA CTAGTAATCA TAATTGTGAGG TAATAGGGCCA CTTCGAGATG 960  
CAGAGGATTC CATTTCCAGAT GTACATCACG GCTTGTGCTC TTAGTTTTC CAACTTGGGA 1020  
CCTGATAGGA GCANAATCTC TCATTTCTCC AGGTCCCAAGG CAGAGATCCT GAAGAATAG 1080  
GCTATTATGT CCTGCTCTCC TTGTGACTG CTTCTTTGCTG CACGGCTCC TGACCCCAAC 1140  
CCTTTGGGGC ACACCTGGC ACTGCCAGAG TAGCTCAACC AAGCAATTGT GCTGAGATG 1200  
GCACCTGGTG AAGCCTGCT GTGTGCCAGG CTTTGTGCTG AGTGTGTGTA CATGTATTAG 1260  
TTCTTTTACT GCTGACCACA TTGTACCCAT TTCAACAGGA AGGAGCAGAG AATTAAAGTG 1320  
GCTTGTCTAA GGTCAATCAG TTAGTAAGTG GCAGAACAGG GACTTGAACC AAGCCCTCTG 1380  
CTCTGAAGAC CCGGTCTCTGA ATTCTTTCAC TAGAGCTTCC TCATCAGGTT ACCGAGAGT 1440  
GGGTCCCATC CACCATCCAG GTGTGCTGG ATGTATGTTT TCACCCCTCG AGGTGTACGC 1500  
TGTGAANAAT TTGGAGACAC TCGTTTATTA TAAATATGAA TATATTCTAA AAAAAAAAAA 1560  
AAAAAAAAA 1569

(2) INFORMATION FOR SEQ ID NO: 46:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1924 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GGGCCCCCCC MCGMKTTTTT TTTTTTTTTT TTAAATTAGG ATATGCTTT TATTAACGAG 60  
AATGAAGAAT TCATTCCTCC TTCCTACTCT TCCTGTGGT TTCTGGACA CAGCTCACT 120  
GATCTGCTTA GAAAGGTGT CAGTCTGCTT GTGGCTTCCC TCTTGTATG ACTGACGCTG 180  
TGTGATGCT TGAGAGATAT CTATCCACTT CATGTGATG AGCACTCCAA TATGACCCA 240  
CATCAATCAT TCTTACTCTA AGATATATRA GAAAAGTTA ATATAANGA CAGGGTATA 300  
AATAAAGGT TTGAANAATC TAGTCACTT CAAAATTTAA AGATTAANA TCCAGAGATA 360  
AAGATTGGGG GTAAATTACA GCTAATAAAA ATAGAGAGAA ACTTCATGTT GGGGGGGGAA 420  
TCTAANAATA TTCTTACATA AATTAATAGT ACACCTGAT TAGATGAAA ACTGTATTTT 480  
CTTTAANAAT TAAGAGCTG ACTCTCAGTT TCACCACTCT GAGCAGAGT TTGACTGCAA 540

CCCAAAATAT ACTATCCCTT ATGTGAAGGT ATGTGACAC GTTGAACCTCA CCAATGAGT 600  
TTTACATCA GCTCTTTTTT CATATGANG CACATAGCCT GCTCCCAATT CAGATATGTC 660  
TTCCATTGTC AGCAGGCTG ACCACCTTCA GCAGGAGTCC TCCAAGATG CCAACTCCC 720  
CTTCCACAG TACACAGCC TTATGTTGTT GTCTGCAAT CCTTGTATT TACCTCATTC 780  
TTTCCCATCT AGTCTCTCAC TGAGTTTAAA AGTTAGGGCT GGAANAAGCTA TGCCTTACTG 840  
GGACAGCAG GAACCAATTT TTTTCTGAGG GAGAAGACAT TCACCTTCAC TATATGCTG 900  
GAGGGCAC AGTGACAAA ACAAGATCA GCTTTCATTC AAGTTCOAGG TTTTCTTCTC 960  
TCCCTGAATG ATTACTGCA AGGTATATG AAGTAAGAGT TCCCTGTTC ACATGACCA 1020  
TCCATAAGG ATACTATATC GTTTTGCAIT CTTCGCCCA TTCTCCACAT TGTCTATCT 1080  
TAAGTCCOAG CCTTTTTCAC TCTCAAAAA AAAAAAAAAA TATTTTTTTC AGCACTGGTG 1140  
TTCAAAAGCA AGCTTTTAT GCTTAATGCT TTACAGCA CTTGTGAGT TTCAGTTGA 1200  
GTCTTAAAA TTCCCATCA TTATCTAGCA GCATGACAG ATGATTAGGA GCAGTCAAAAT 1260  
CCTCTGAAT CTTTCCTTAA TAGGCAGCCA TTGTGAAACT GCATGAGCTG ACATCACTAA 1320  
ACATTATCA GCTAAGCCA AAACCAATA AAGGCCAGA CCAACATCTT GGTCTCTTAA 1380  
AAGCTGCCA AAATCTTAA GTCAAGCCA GTAAATGCCG GACTGTGGAT CATGTCACTG 1440  
CAGCTGACAA TGAATTAACA TAGAGACAT GCNAACCCCA TTAAAGTTAA AAGTCCAAA 1500  
CTAGTCAAC GCATCTCTT ATTGGGAAA AGTGAGACTA TTATGCATC TTGTGAGGTT 1560  
TGCACCTTG CATGAGNCC ACCATTTGCA TTTCTTTCTT CTTTCAAGAA GCACCGGTAT 1620  
CTGTTCCAG GGCCTAACAG TACGAAAATA CATCTGCGA TCACACTCTT GAACCCAGA 1680  
CTGTTCTCAT TAAAAATAT TTGTGTTGT AACAAATTA TGAATATCA TCCAGACCC 1740  
TGGTATAGC ATTATTAAGT AAACCCCTTA ATTCCAGCT TTTTGTGTTT TTTAAAAAA 1800  
CCCACTCAC TAAGATTCAC AATTCATTC TACATACAA TTAAAGCTAG TAAGACACA 1860  
CTAAGTCA AGTTTCTCA TTCTAAGTG CAAGAAGCTA ATCATCTGA AGTGACAGG 1920  
GTAA 1924

(2) INFORMATION FOR SEQ ID NO: 47:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 475 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

60  
120  
180  
240  
300  
360  
420  
475

TCCTCTGCGG CCGGAAAGC AAGGACACAG TGAAGAAAC CCGAGAGACT TGTATCCGCC  
AGGAAAGCCA TTGCGGATTC TGAGCCCTTG AAGGCGAAG AGGAAAGAG TTATTACAGA  
GCCGATNAG AACCTGCTTC ATGAAAGAGG GCGGACCTTG TAAAGAGACT CATTACTAAC  
AGAACTGCG TGCGAAATG CTGCTGCTC TCTTTGAGAA AACCAACAG ATACACTGCG  
TGGAGACCA GGTGGGACA GAGAGGCTG GAGAGGAT CTGGAGAG AGATTGAGAA  
TAAACAGAC AGCTTATTC ATTCTGTAA ATTCTATAT AAGAGCTCA GTGTAGAAA  
TAAAGTTAT CTACTAGTGG GAGTTAGT GTTCTGTTT GTTCTGCTT CCGTTACAA  
TAAATTAAT CCGTTGGA CTACAGAGT ATGCTCTCC TTGACATAT GTGAA

20 (2) INFORMATION FOR SEQ ID NO: 48:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 346 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

60  
120  
180  
240  
300  
346

AAGGAGAG GACTGGAT CAGATCTAT TTACAAATG AAGACCGAAT GCGAAAGTC  
ATGCTGAAA TTCTGAGCTT ACTGCTGCG CTGCTGGAG CTGCTGAGG TGAAGAGAG  
GAGGAAAGG ACTTATACAG GAGGCGAATG AAGTCACTCC ATGAGTTTCC TGAACCTGCG  
CGAGTACAG ATTACGTTT GAGCTGAC GTAGAGACT GTCCAGATGG CTACTTGCTG  
GCGAGATGA AAGACAAAGC CAGAGCGAAG CCGGAGGCT GTTAAAGAG GCGAGATCTT  
GCGGAGCGA CTCTGTAG ATCTGCTTC CTTCAGAGC TGAACC

45 (2) INFORMATION FOR SEQ ID NO: 49:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1366 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

60  
120  
180

TAGGCTACG CCGCAGCCC CCCCCCAAT GCAAGTTAC TGGCGAGGT AGTGGCAGC  
TTTAAACAA GCTGATATT CAGGCTGCT GGAAGCTAC TGAAGATGA TACCCCCCA  
GTAAAGAGC CAGGTGATCT TAACTCTTT CAAGAGCAG GCTCTCTGG GAAAGCATGT

240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1320  
1366

CGTCAAGAG CAGAGCAAC CCTCTGAAA TGGATACAAA ACTGACTCT CAGCGAGGCA  
GCGGAGCTT CTATTGTAC AGTAAAGCA GTTATGCTGG ATGATACAT CAGCTGGATT  
TTTGGTTTAA TGATGTAGAG TGTCTGAC TAAAGGCTTT TGAGGCTCC TTGCGCTGCG  
AGTGGTGA CTTCCAGAC CGAATTGCT CAGTTGCTCT GAGCTGATG TGTCTCTGCG  
TGGTGGGCA GCGCTTGCA TGGGAAAGGA GCTCTGCG GCGGAGCA GCTGGGGGTG  
CTGAGCTAG GCGAATGAA GTTATGAGG GAGCTGTTT TGATGTAGT GAGGTGATC  
TTCAATGCCA GCGGAAAGTC GCGGAGTGA GCGAGGTGCT GCGAGGCTTC TGTCTGAGC  
TTGCTCTCT GTGCGAGGGG AGCTGAAA AAGTGTAGG GCGCGCTAC TGCAGAGCC  
TGTCTGCTG CTTCTCCAG CAGTGTCTG GGGGTGATT CCGTCAAGCT AAGTGTGAA  
GCGCTTACT TTCTCCAGC AAGGAGTGG CAGCGAGCT AGCTGAGCT TGCAGCTGCG  
AGAAATTTCA CTAGCAGGT CTAGGCAAGG ACTGGGCAAT TGAAGAGAG AAGCGAGCT  
GTAGATCTGA CTTGAGGCG GCGCGCTTCA CTTGCGCTG GCTGTGCTG GTCTTAAAGT  
TTTGTCAAGT TGTCTGTTT TGAATCTTC AACTAGTGA TAAAGACTGG AAGGGGATGA  
CGCGCTTGG AGGTGTTCT TTAAGCTGAT CCAATATATA GCGCGCTGG ATGCTGTGAG  
AAGTAAAGCA GAGTATGCT GTTTAAAGC CAGAGCTGG GAGCAGGCTT CTTACAGCTA  
ATTCTCTTC CTGTAGACTG AACAAAGTTC TAAATTAAGT TAAAGAAAG AACAGCTGTC  
GTACAGCAA GTTCTAGAG CAGCTCTTTC AGTTTCCCT GTTACAAAT GCTCTCAGAT  
TCTTATGAA GCAAGAGCC TTAGGGGGGG TGCGCAGAG AACAGAGAT GTTAAAGCTG  
AGCTGTGAC AACAGGGGCT TGTGTGCTG TCTGTGTTCT CTGCTGCGG AACCTTCTCA  
ATTAAGCTTA TTCTTATTT ATTAAGAAAA AAAAAA

45 (2) INFORMATION FOR SEQ ID NO: 50:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1405 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

60  
120  
180

GAGTAATTC CTGTAGCA CTGACACAC CAAGAAAGT TTAATTTCC CCGCAATTC  
ATGATTTTA CTTCTGTAC AAGAGAAAT TTGCTGATAG GTTGTGGCT CCGCAGTTC  
CAATTATGC TAAATGATTC CATTCTAGG CCGAGGCTTC TCTTACAGA TGTGTTTCC

TTTAACTCTT GGCATGATA ATAGATTTT GGTGATGAA AGAACCCAA TAGGCAGAT 240  
AGTCCGCCCA GGCCTGATA TCCATAAAG GCTTGGAAAT GCATATATGA ATGTGCTTTA 300  
5 GTCTTTTCTT TGTATTGAA AAAAAAACA AGATGGGCTC AGATGATGC CTAGCTAATA 360  
ATGCTTCTTA GCTGTGTACT CATATCTTTT CTTTGAAATG AGTAGTGAA GGAAGGAGA 420  
GCAAGGAAA TTAAATGTCC TTCTAGTATT CTCTGACTC AAGTCTGACA TATGAGATAA 480  
TAACTATAT TGAATGCGA AGAATGTAT CTGAACAAG AGAACAGTTT GACACATTTA 540  
TCATGCTTTC ATATTACATA TTAACTGAAA CCAATTATA ACHATATGAA ATATCATTTG 600  
15 CACAAGCAA AGCAACTTA ACCTTTGTTT TCTTTTCTA CATAGCAGAA ATTGATTTT 660  
TTTTTATTTT TTAGGGGAA CCTATATAT ATATGACCCAG TGAATGCTTT TGGTGACTTA 720  
ACCTTATGAA TTCAGGTTAC AATTGATGTT ATCTTAGATG GTTACTACTT TGAAGAAGAT 780  
840 GTTGTGCTCT TATGTGACAC GAGCCAGAGC CTGCTGGGGA ATAAACAAG CAGGTTTCAT  
GCCAACCA ACTGTAGCT TTAGTGGGCA GATGGGAGT GGTTCACAGA CTTCCCAAAA 900  
25 TGTGGGGCTT TGGGATTTT CCAACACATC CAGGTGTGT TGTTCATCT TCTCTTTTC 960  
ACACTCTTGG ATGGATWATT TGBAATGTT YKBAATWICY YKBAATWICY CCAATAGCC 1020  
1080 WTGRCOACC ATTCTTWTAT ACACATAC CAATATGATC ONTATGTTG AATATATAGA  
1140 AACCTGTAC CAGCTTATGA KTHACCONHA WTTTTCCTT GTTTGTGAA TTGATATGA  
1200 AATAGCAGGG CTAGGAATT ACTGCAAGT TTTAGCCTGT GGTATATACC TTAGGGTAT  
1260 TTAATATATT GTATTTTAT TTAAATGTC ATGAATGTTT GAAGGAAACA AATATATCAG  
1320 GAGTGGCTCT TTGCCATGGG TCTTATTTTC ACCCTCTTTT CTGTAGGAAA AAGAACAT  
1380 GTCTTAATCT ATTTTAAAG TTTTGGTAT AGTTTCTAAT TCCATTTTA ATAAAGTTT 1405  
TWTATAAAA AAAAAAAAAA AAAA

(2) INFORMATION FOR SEQ ID NO: 51:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 504 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

CGGATTTTCT AGGACCCCAA AAAAAAAAA AGGNAAAAA AAACCCNCA AACCCNCA 60  
AATCCAAAA AAAAAAAAAA TCCAAAAA CAATAAACT ATAAAAAGA AAGATATA 120  
60 AACTTCAGA GAATTAAT TTAATTAAT AACTTACGA TTAATTAAT AATATATAT 180

TCACCTAGCA ACATATCTCT GCGTCTCTC CTGCTCTCAT AATCAAGACA TAGCGATTC 240  
10 TCTGCCCGGG CCCCTTCTG ATGCTCTCTCC GGGTCTGCTT CCGGCTGCGG TCTCTGGGGA 300  
CCCTCCAGAG GTGAGAGTGG GCTGATGGCC TGGGTCTCTG GTGGTTGATG GTTTTCTCTC 360  
CCCTACCTTT TTTTCTTGG TTATATCTGA TTGATTTTCT TTCTTCTGTT CTGATTAAC 420  
CACCCTCTGG GCACAGGATA ATAAACATG TAATATTTT AAGAAGAAA AAAAAAAAAA 480  
AAAAAAGTGG GGGGGGGCCC CGAA 504

(2) INFORMATION FOR SEQ ID NO: 52:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 777 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

NAAGTATCTT GGCCAGTTTA TTACAGAGCA CGATTAATGA TTCCATGTGG ATAGGCGATA 60  
120 ACNTACAGAG AATGAGACTA TCCAGAAAT GGGAGAGGC ATTGAAACA ACATGAGTAT  
180 CTCAGGACA GATGATTTCA TTCTGCTATT GGTAGGCTTG GAAGCANGG TCAGAAATAG  
240 CAATAATGG ATACCAAAAG CACTATTGCT CAGCCAGCT AGTGGAAITA GCTGCCCCAG  
300 TAGGAGAAAT GCAGTTTTTG CTCTACACTA AGTTCTCCAA CTCTTGATTA GCCTCCNAAA  
360 ACAATGTTA GGGGAAAAA AGGCAGCTGG TTATGAAAG ATATATCTCA TTTCATTAAG  
420 AATCAATGT CAATGCTGTT AATGAATCC TTTTATCTTC AGGACAGAG CAATGCCCTA  
480 AACAAACCC AGCTCAGAG CCTGTATGC CAACCTAGAG GGTACCCAAA CACAAACTTA  
540 GCATAGAGGT AAGATCTCT ATGTCTTTTG GTGAGGCAA AGCCATTTGG TTGTACTTTC  
600 AAGAGAAAT CTTTCTACCA AGTCTTCATC ATATGTTATG TGGCAGAGT CTCAGATTGT  
660 TTGCACCAT GTGTATAGC TGAAGATAG CTGAAAGGTT AGTTTGTATC CTGAAACCT  
720 ATTTACAAAT GCCAGCTGAT GTCCCTGCTG CCACTTAAAA AAGGCTTGG TCTGGCATAG  
777 GCGAAGAGC CTGTGTCTCC CTGTGCGCA TTCTNGGCTC GAGGCCAATT NCTTAT

(2) INFORMATION FOR SEQ ID NO: 53:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 602 base pairs  
(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

5 ATGACATGAG TGTATACCC TCGATCTT GAGGTGGG ATGAAACAT GCTGTATCA 60  
CTCTGTGAC GGTATTAATC CATTATCCA CAAAAGACA CATCCATCA TCAACATATA 120  
CATGTGTGG GATGACGAG TCAATGTT TGAAGGGAG TTGTCTCT TTTTCTTCT 180  
CATTAATCT TTAATTTGT GTGAGTATC AAACAAACA ACAAAAAA TGTGTGAAA 240  
AACCTTCAT AGCCTTTTC TATCAATGC TTAAAAATAT AGCTAAATA CACACATCT 300  
GCGATTTTT TCTTACATG ACAGATCTT TACCTGCCAT TTAAATTTAG CCTGTATTT 360  
TTCTACATA TATTTACATG TACATGTAT TTGTATTTTA AACAGAAA AAACATTTA 420  
AAAAAGAAA AATTACTGT AGCCTTCAT TATACTATA TATTTATAT ATTATTTGA 480  
CATTTGCAA TACTGTGAA GTTTTATCT TTGCATATAC TTAAATGGA AGTATTTACC 540  
CTTAAAAAT CAAAAATATA TTTTCAAGG TTCCGTTTT GTGTGTGAA AGAGTAAAT 600  
GA 602

30 (2) INFORMATION FOR SEQ ID NO: 54:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1749 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

40 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

AGTCACTGAC TTGAGCCCG TCGGGGGAG TTCCGCCCC AGAGCTGGT GATGGGAATG 60  
CCTCACTTCA GTTTGAAAG GATCCGATC CAAGGGGTT AAACGAGCG AACCCGATC 120  
CCGACACAAA CTTCGCGGCT CCGTAAAGG CACACCCCG TACGATGGG GAGCCGAGC 180  
CACTGTGCA AAGTGTGTG GTGTGGGAG GCGGCAATG GCGAAGATC GCTGTGTGAG 240  
GATTAATCCC AGAGACGTT CAGCAAAAC TACAGATCA CCGTGGAGT GGAATTTGCT 300  
CTGAAGTTC TCGAGTGTG TGAATGAGG AATAGTGGCG TTCAAGTGT GGAATATGCA 360  
GCGCAGGAG GCTTCACTC TATGACAGA TTGTATATC GAGATGCTC TCCCTGTCT 420  
ATTATGTTG AGCTTACAA TGCACATAC TTCAAGACA GCGAGAGTG GAAACAGAC 480  
CTAGACAGA AGCTCAACT ACCAATGGA GAGCCGCGC CCGCTCGCT CTTCGCAAC 540  
AAGTGTATC TGTCCCTTG GCGATGAGC GCGACACAGA TTGACCGTT CAGTAAAGG 600

AACGTTTCA CAGTTGGAC AAGAACATCA GTCAAGAGA ACAAATAT TATAGAGCT 660  
ATGAGATCC TCAATGAAA GATGATGAG AATTCAGAG AAGATATCAT GTCTTTGCC 720  
ACCGAAGGG ACTACATCAA TCGAAGACC AAGTCTCCA GCTGTGCTG CTGTAGTAG 780  
TGTGTGCTT AATTTCATC CAGTTCAG GAGCTTTTT AATGATCTTC CTTTGTGTG 840

5 CCCACTGAC CATTATTTA AGTACATTT AATTGTCC TACATCTGT CCGATTAAGA 900  
GGCCCATTT TCACTTAAA AAAGACATG GAGCCGAGT GGAATTTGCG AGTCTGTGA 960  
TTAGCTTTC ACATGTGCT GTCTCAAT TATGCGAAT TATGCTTGG GTATAGATC 1020  
TTCTGATGAG CCGTCATTT GTATCCGGA AATTGTGAG AAGATTTGA AATCAGACC 1080  
TCCGTTTTAG AGATCATAT TCTCACATC TTCTGAGCT AATTTCAT TTAGATTTCA 1140  
TTGATATCAT GACTTCAT TGAAGGAAA ATGAGATCA ATGTCAATTC CCAATTTCT 1200  
TGTAGCCGT TGTTCAGAT TCTTTCTGC TTGAAATGA AACATGAT TCTGGAATCC 1260  
AGAAAGAGA GTCTGGCAT CTGTGATTT TTGCATATCA GATGTGTCC AAGATCATC 1320  
GTATTTTGA AACTTCATC GTCAATATA AGTTCCTCT GTCTTGCAT CAAAGATAT 1380  
CAAGTTTTT GCGCGGCAT GATGCTCAT GCGATATC CCAACATTT GGAAGCCAA 1440  
GCGAGCGCA TGCATATGAG CCAAGATTC GAGACGAGC TGTGAGCAT GCGAAAACC 1500  
CGTCTACT AAAATGACA AATTAGCA GCGGTATGAG CAGTGTCTG TAATCCGAG 1560  
TACTTGAG AGTATGTTG GAGATGCT TGAATCTGG AGCCAGAT TCGATGAGC 1620  
CGAGATCTG CGACCGACT TGAAGCTGG TACAGAGAA GAGATCTGTC TCAAAAAA 1680  
AAAAAAA AAAATCGAG GGGGGCCCG GTACCCAAAT CCGCTGATA GTATCTTAT 1740  
ACATCTAAA 1749

40 (2) INFORMATION FOR SEQ ID NO: 55:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1896 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

AAAGATGAG GCTCTTAT TTCTGAAA ACQAATTTG AGTATCAT TTTCATTA 60  
CATTAATTT CTTCAGTGA ACTACATAT GTTCATATCT GCTTATCAG GACTATGCG 120  
CTCTGTGCT ACTGATCCA AATGACAT GTACATTTA CCGCTCGCT TTGTCTTAT 180  
GGTGTGCTG TGTATATG AAAATGACA GGTGTGAG GGTGTGAGC ACAGATGTT 240

5 TTTATAGTCT AGACCTTTA AAAACCCAG CAGATGTAA TTGATATTT GTTATTCGC 300  
TTTTTTTGA CAGATGTTG AATTAAATG AATTGAAGG GAACCTGAG GTACTAGAC 360  
GTTTATTAA AGAAAAA TGTCTTCCA TTGCTGTAA TCACAAGG AGAATAAC 420  
TTGTTTCCTT GATCTGAC AGCTCAGCT AACCTGGGC GAGCTGTAT TATTATAT 480  
ATAATAGTAG TAGGAAGTTA ATATGCTT CTTCTGTTC CAAGCACT ATTACAATT 540  
CTTTGAACC GTAAATATCA GAATGAATCC TCTTCCAGG GATGAACA GAAGTTAAT 600  
GTTTACAGT GTTTGAAATT GTGATCTGA ATACACAAA ATTAAACCA TGAATTCCT 660  
AATTTTCCA CTAGAGGAG AGAACTTGT GGAAGGTT TTTTTTTC TTTTTTTTT 720  
CTTAAGAAG GGCAGCAG GTAGTAACCT AAAAATAGT CCCAGGATA TGAGATGCT 780  
CCTACAGGT TAAGAACAC ACTGTCCAC TGTATGGCT TGGCCCTGAG TGGCAGGGA 840  
GCTCACTTG ACCCTGCCAT GTTGGTTGA CTTACTAAGA CACAGATC ATTGTTTCC 900  
TTGACAGGG TCTCAGACC TCGAGGATG TTAACTAAGA GAAGAACT CTTTCTGAA 960  
TATTGACATG TAAGAGACA AAGTAATTT TCTGAATTC TGCATTTG AGAATCTCC 1020  
AAGCAATTGA CAGTGATTTT AGCTCTGTC AGCAATTTTC CATGAGACT TTCATCAAT 1080  
TGACTCTTTA GTTCAGAGT TCCCATGAT TGTGAGCAG AATATTATCT CTTTAGCCCT 1140  
TGGGATCCA GCTGAGACA ATCTCTTGA TTTTTTACC GGTGATGTA CAGATATCAT 1200  
TCTTGTTGA TGCATGACT TGAAGAGTT TGGGAGCTC TTTAGCATA TCAGCTAATA 1260  
GGAATGAAA TCACAGTGA TAGCAGTGT CATTCAGTAA TTTCCTACAA GCAGCACCC 1320  
AAAGAAATA TAGTCTTAAT CTTTACTATC CACTTCTAAA TTTAAATGTA ATTTCATCA 1380  
TGTATATAGT TGTTTTCTTT ATAAATTTAT AAAAATTTT CATCGAGCT TTAACTTCCA 1440  
CTTCCATGCT ATCGATGTC TTGGCTCCA TGCAGAACT TCGAAGAAA ACAGGCAGA 1500  
ATCATTTGC ATAAATGACC AGATCATCAT TTCTGCAAC TGAGAAATAT ATTTCATCAT 1560  
TGTCTCTAGA AGCTGCAT TCTTTACTTT TCTTTGCTGC ATTATATCT AGCTGCATC 1620  
ACTGGATAAT GTGGATGAC TAGAAGTTC AVATATCACT GTAGGTACA GTTAGGGSTA 1680  
ACACTTTAGA GGTATATAT TTTTAAAAA CTTTCTTGA ACTCTGGGC CAACATGGT 1740  
GAACCCCGT CTTCTTACTT AAAATACCC AAATTAAGC CAGGGGCTG GATGGGTGG 1800  
GTGCTGTGTA ATCTTCAGCT ACTTNGGGA GGGCTTGAG CAGGGAGA ACTGCCCTGG 1860  
AKCCCCGGG NCGCCAGNA GGTTTGCCAG TTGACT 1896

## (2) INFORMATION FOR SEQ ID NO: 56:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1753 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

5 TCTTTTAAA ATAGACATTT GTGGGCTCA CACATATAT GAATAGTAC CCTTAHAAA 60  
AGAGAAAAA AAAATCAGGC GGTCAAATTT AGAGCAACAT TGTCTTATTA AAGCATAGTT 120  
TATTTCACTA GAAAAATTT AATATCAAGS ACTATTACAT ACTTCATPAC TAGGAAGTTC 180  
TTTTTAAAT GACACTTAAA ACATCTACTG AAACCTTGAT CCAATCACA CCTGTATTAT 240  
TTTCTTAAA CATCTGGAA GCTTAAGCTT CTGAGATCA TGTGGCAAGT GTGATGGCCA 300  
GTAAATATCC AGAGAGATG TTTAGTAGCA ATTAAGGCT GTTTGACCT TTAAGAGACA 360  
GCTGGGCTGT AATGATTCCT GGGGCCAGAG TGGCATATG TTTTACAAA ATATATGAT 420  
ATGTACATG TTTGCAATGT TGTGTGCTG TTGAATTTT GAAGGCCAG TTGACCAATC 480  
ATAGAAATTA TTAATTTCTT TCATATGGTT TTTGGTTGAC TGGCTTAAAG GGTTTCTCAG 540  
AATATCTATG GGCAGAGCAG CATACAGTTT TCCATCTTAA TAGAATGAA ATTATTTTG 600  
TATCTACTGA TAACAGATC TGGTCAAT GAAAAAAT CATTTTATCC GTCTTTTAG 660  
TATATGTTTA AATTAATAT TATGTGTCT GCATATGCA GAAAGCTCT GAGAGACACA 720  
GTTTCCCATT AACTTTTCT GACCAATAGT CTTGGACAG TGTCTTCTC TTTGGAGAGA 780  
GGAAGGGGTG TGTGACATG GCTAACATC TTCAAATACC CAATTTGTA TAGCATTAAT 840  
AAAGTATTTA TTTATGCTT CAGTATATTA TTAATTAAT TTTTAGSTAA TGCCTATCTC 900  
TTGCTCTATT AAGGAAGAA GCATCACTA GAGATTCAG GATAGTTTG TTTAAATTTCT 960  
TCCAGATPAC ATGTTTTTAC AGTGGCTGC TATTAAGGA AGTATTTCTT CVATACACT 1020  
TGTTTTAAAC TTTGAGACA TTAGACAAA TTATGCAATG GTTTGTGAG ATACGAGCTT 1080  
GATGGTCTG TTTAATCAGT TTGCTTCCA AGTGGCTAC TGAAGAGCC CTAGAGCTGG 1140  
TAGAATTAAT AAGATTTTCA AAACTTTTCT ATTCTTTCT TAAACCTACC AGCAACTAG 1200  
GATTTGATA GCATGAATG GTATGATGA GAAATTTGA CCAATTTGT TTTTGTGTG 1260  
TTGTGTGTGT TTTGAATTTG AATCATCTT TATTCCTTTT AAGATGTATT ATGTATGAGT 1320  
GTGAAGATCC TAGGAACTT ATGCTCAGAT ATTCATGTA AGTCTCCTT CACCTGTATC 1380  
AGAGTTTCTAG ATGCTCACT GATAGTATGT ATTCTTTAG TAAGATGTG TTAATAATTAC 1440  
AATGATCTTT TAAAGATG ATGAGTTCT GTATTTATG TGTGTGTCT GGTCTTAAGT 1500



GGAGCCAAAT AAACAAGTTT GATATGATTT TTTCAGATGT TGAATCTGAC ACACATGACT 1560  
TTGAAAATTTT CATTGCAATCC TGAATAGACA ATTAGAAGAGC CAAATATATTT TCCCTCTTAA 1620  
TCCCTGAGAT TTCACTACTT TTATGTATTA AGTTGTATTT AATTTATAAA ATCTGTGAAA 1680  
GAATAAAAAG TGAATTTAAA TTAAAAAAA AAAAAAAA AAAAAAAAAA AAAAAAAA 1740  
AAAAAAGAG GAG 1753

(2) INFORMATION FOR SEQ ID NO: 57:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1220 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

CGGAGAAATTA CTCAGAGCCC GGTGTGTGTC TGTGGGAGAG GAGAAAGAGT TTGTAAACCC 60  
CGAGAGAGAG TTCTGCTTAC CCGAGGCCCC TCCCTGTGCG AGAGCCCCCG GTGAGCCGAC 120  
GCTGATCAGT TCTGACAGAG AGCGAAAAC TTCACTAGAG GACTGTGGGG ATAAAGAGAGA 180  
AGGTGAAATTT ATTAACTCA AGGTGATGAG ACAGATAGAC AGTAGAATTC ACTTCAAAGT 240  
GAAATAGACA ACACATCTCA AGAACTCAA AGAATCATAC TGTCAAGAGC AAGGTGTGTC 300  
AATGAAATCA CTCAGAGTTTC TCTTTGAGGG TCGAGAGATT GCTGATATAC ATATCTGAAA 360  
AGAACTGGGA ATCGAGAGAG AGAGTGTGAT TGAAGTTTAT GAGAAACAAA CGGGGGGTCA 420  
TTCAAGAGAT TAGAATATCT TTTTATTTT TTTCCTTTC CTCATGCTT TTTTATTTT 480  
AAAAATAGTT CTTTGTAAAT GTGCTGTCA AAAAGGAAAT GAAACATGAC ACCCGATGTC 540  
TTTGAAACAT CTTGTATTTT GAAATTCAGT GCTCATATAT GATTAATGTT TGTTTTCATT 600  
GTGCTGATTT TTGTGTATCA AGCTTCAGAT CCCCAGATAT TACCTCTTCC TTTTAAAAA 660  
TTAGCTGTGAC ACAGAGAGGT GACCTTTTC AGACATTCGC ATTTTCAGGC TTGTGCTGAT 720  
AAATAGAGAT GACGAATGCA AGTGTTCATA ATGCACTTTC AATTGGCCCT GATGTGCTAG 780  
CAATGTATTA CTCACATGCT GAGCTGTGAC TTTCAGTGGG AGATGAGAT TTTCAGAGAA 840  
ACTGAACTGT GGAATAATGA CTTTCTCTA ACTTGAAAGT ACTTTTAAA TTTCAGAGTC 900  
TGGAGCAAAA GAAAGAGAGAT ATCAGGTGGA AGTCAGAGAG ACAGATAGAG TGAAGTAAAT 960  
GACTACTCC AAAGATGGCT TCACTGAGAG AAAGGCAATTT TAGATTTTAT TAAAAATCT 1020  
GTGAGAGAGT CCGAGAAAAG TTCTAAATTT GATTAAGAT TAAATAGCT ATACATGAG 1080  
AAATGAAATC AACAGAGAG TCCCTCTTTT GATTTATTTT GATCTTTTGG GCTGTGAGATA 1140

TGGGTTTAAA ATGAGCAATG TCTGTAGAG CTTGATTTAA ATTAACATAA TTGTAAAAA 1200  
TGTAAAAAAA AAAAAAAA 1220

(2) INFORMATION FOR SEQ ID NO: 58:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1049 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

TGGGCTCTCC AAGACAGACA TCTACTAGAC GTGGGTGACC TCTGTAGACA TCACTGAACTG 60  
GAGGCTCTCC TGAATTTCTG GTGCAGCCCA TCAAGAGACC ACAGGCGCTG GAGAGATGAT 120  
GGGAGATCTG GCGAATGAGG AATGTGTACA GAGAGAGAG CCGGAGATGA GACACACTAG 180  
CGAGCCACA AAGATAGACA TTCTGTACA GAGCTTCTTC AATAGAGGCC ATGTGCTCCC 240  
CCGAGGGGAT CTTGCCCCC GCGAGAGACA GCGAGTGC ACCTGTGGTG CTGCTGAGTC 300  
CCCTTCATAT GACTGACAC GCGAGCTGAT GAAATGAGC TTTCGCGAGT GCGATCTGAG 360  
GAGCATGCT GTGAGAGCCG TGAATCTCAT CTTGCTCTTC TTCTGCTCA TGAATCTTGG 420  
TGTGTGTGAC CTCCTCTGAG TTGGCTCTGT GTACGTGATG TCCAGCTGTA GTCAAGGATG 480  
ACCTGTAGAG GCTGATAGAGG GTGGGTTTGT TGAAGAGAC TTGCTGAGCC TTGGTGTGAG 540  
AGCAGAGATA TTGAGAGGGG ATCTGTGAT GCTTTAGAG TATGATACAA GAGGGAGACA 600  
GAGGTGTGAG TTTCCTCTTT GTGTTAGAG TGAAGCAGAG GAGAGCTTGA GTCCAGACAT 660  
TTCCCAAGAT GTGGATGAT CCGTGTGATC CCGAATATCT TTTAGGTGAT ATGAGGCGCTG 720  
CATTAAGTGA GACAAATACA GAGCAAGAAA GCAATGCCCT TCCCAATCT CTCATGCTT 780  
TTATGCCAG AGAATGTAG GTGATGCCA ACATGTTCAG ATGCTGTGAG AAGCATGCC 840  
GAGGTCTCT CTGCTTAGAG ACAGAGACTT GGGCTTAGAG CAGGTGAGAA AAATTCAGAA 900  
CTTTTTCAG ACTGTTTTGG TTGTAAAGAT AATTTTAT TGCATCTT ATTTGTTAG 960  
ACAAGTGTGA GTGCATCTT ATTTATGTG ACCTTTTCAA TAAATAGAT TAAATAGAAA 1020  
AAAAAAA AAAACTGAG GGGGGGCC 1049

(2) INFORMATION FOR SEQ ID NO: 59:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1776 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAGAGGATG TGAAGCTAGA GGTCCCGCAT GCTGTGTCGG ATGGGAGCA CAAGGCTGAG 60  
10 GGACTGGATT GTAAAGGCAC TAAGTCGTTT TCGGGTGAGA ATCAGACATG GGGGACTCTT 120  
AGCTTCACAT CTTCTTTCTT TCGAGTCTTG GACATCTCTG GCCAAGTCC CCCACATCTA 180  
15 GTCCAGTCAT GAGTGCAGAA GTGAAGTGA CAGGGCAGAA CCAGAGCAAA TTTTCTCTCC 240  
TAGCCAGTTC GGGCAGGGGG GGAAGGCTGG CCACACTCAT CCATCAGTGG CTGGAGGCCC 300  
CTGGTGTCTA CTTGTTTGA GAATCTCTGG ACATGCCCAA TGTTAGAGAG CTGGCTGAGA 360  
20 GTGACTTTTC CTCTACCTTC CGGCTGCTCA CAGTGTCTTC TTATGGGACA TAGGCTGACT 420  
ACTTAGCTGA AGCCCGGAAT CTTCTTCAC TAACAGAGGC TCAGAGGAT AGCTTTGAC 480  
25 ACCTTCAGT TGTCAACCTG GCTCTAAG TAAGTGTAT OCCATATCCA GTTGTCTGG 540  
AGGCTCTTTC CTTGCTAAT GTGGGAGC TGGAGACT TGTGATTGAG GCTGTGATG 600  
CTAGCTGCT TGTGTGCTTC CTGGACAGC GCAACAGGG GGTGAGGTT GACTACAGCA 660  
30 TCGGGCGGGA CATTGAGGC CAGAGCTCA GTGCCATTGC CCGAACCTT ANNNANACC 720  
ATTAAAGTTA CGAGGACAG AGCAGCGCA GCCACATCTC AGGACCTTCA GCAACACTG 780  
ACTGAGCTGA GGGAGCAGC TCTTGGACC AACGAGGCC AGCGAGAG AAGGCTTCAA 840  
35 AGGCGAGGG CTTCCGAGG ANGCCAGGA TTTGTCTCAA GTCGATTGA AAGTACTGTC 900  
GTTTCTCTCC TGGGATGTG GAGTCCGAGC TGCCTGCTG CCTCTTAGCA GTCTCAGAG 960  
40 AGCTTTCTGT GCGGCTGGCC AGCTGTAAT CCTAGTTTCA TGAGCTTTCA CTTCCCTTAA 1020  
CCCAAAACAT AGATCAGACC TTCTCTAGGG AGGAGKCAA TGTAGGTCTAT GTTTTGTGTT 1080  
GTACTTTCTG TTTTGTGTA CTTCAATGTT TCCATTGCTC CCGGCTGCCA TGCTCTCTCC 1140  
45 CTTGTTTCTT TAAGAGCTCA GCATCTGTCC CTGTTCATTA CATGTCAATG AGTAGGTGG 1200  
TAGCCCTGAT GGGGGTCCCT CTGTCTGAG CATACGCCAG AGCGTTTCTT TCTGCCAACC 1260  
50 CATCCCTGCA TGCCCTGATC CCAGTTCTTA TAGCCCTAGC CTGACTTATG GAGCAGCTTC 1320  
TGAAGAGCCA TAGGGCCCCC ACCTTTACTC ACACCCCTGAG AATTCCTGGA GCGATCTCTC 1380  
CATGCCAGGA GTACCTGAGC ATGTTCATCC TAGAATCTTG TCACACTACA GTCAATTTCTT 1440  
55 TTCTCTCTTC TGCGCCCTGG GTCTCTGGA TCTCTCTGCT TCACCCCTGAG AGCCTTAGAA 1500  
TGCCAGCGGT TTCTTAACAT GTTGAGAGAT GATTTCTTCT TGCGCCCTGAG CATCTCGGGA 1560  
60 AGCTTGATGG CAACTCTGGA AGGGTTAAT CTCTTTTCTT GAGTTTGGTG GGGAGGGGAA 1620

GGTATATAG ATTGTATTAA AAAAAAAAAA GTATATATCC ATATATCTAT ATATATATAG 1680  
ACCGAAGAT AAATCTATGA GAATCTATC TACAAAGAAA AAAAAAAAAA AAAAAAAAAA 1740  
5 AGGAATTCGA TATCAGCTT ATCGATACCG TCUACC 1776

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 443 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ACAGATAAT AATAAATAA TAAATTAAT TAAATAAAA ATCTGAGCTA ATCTGAATAA 60  
ATTGAGAGAT TTACATGAA ACCGAGGAT TCTGGCTTCC CAGGACAGT CAGAGAGGCT 120  
25 AGTAGGACAC ACTGTCTGC TTGGCTACTT TCTTTGGAAC AACATGAAT CTAGCTCCTT 180  
TTTTTTTTT TTTTGGGCC ACTTCATCCA TTCAATGAC CTGGCTGGCC TCTGCAGGTA 240  
AGTAGATAG CACAAAAAT GTAGCAGAGG TTTTGTGCTT GAACTAGCTG GTTTCAGGTC 300  
30 CAGCTCTGCC ACTGTGAGC ATGACCTGCT GCGCAATTC NCACAGAAAT TTTTTTTTTT 360  
TTTTTCAGTG CTCAGTCCC CCTATTGGAG AATCTGCCC CCGCTGGGA CAGAAATGTC 420  
35 ACCCTGGCCC CCGAATCCC TGA 443

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2888 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTAATGTTGT CAAATACCAC CAGGCCAAAC AGAATTTATA TGACCTGGAT GAAGATGATG 60  
ATGTATAGC TTGGTTCTT ACTTAACAGA TGAAGTTTGC AGCCTCAGGC GNCCTTCTCC 120  
55 ACCAATGCC TGGGCTAAGC AGTTCCAGC TTTCATATGC CAAGGCCCTC CTTCTCACA 180  
AAGTGTTC GAATGATGA TACAGAGTC CTGCTCTGCC TTCTCTTAT TTGAAACAAAG 240  
CCTTGACAAA CATGTCCGG AGACTGTGA ACNAGAGGA ACCGCCCAA GAGCTGGCAG 300  
60 CTGCTGAGCC TTTTCTCAGC CCATTTGAG GCACCAAGAT GACTGTGAAT AATCTGCACC 360

	CTCCAGTCCAC TGAAGAGAC ATTGTGAGC TTGTTCTGTGT GTGTGGGCC CTGACAGGAG	420
	CTGCACTGTGT CCAATCTGGG GTAGCGAGAG TGTGTGTTGT GAAAAAGAC GATGCCATCA	480
5	CCGATATTGA GAGGTACAGC AACCGGTGTC TGGACGGGCA GCGATGAGAG TCCAGCTTTC	540
	AAGTGAATGG GAGTGTATTC AACTGACAGC AACCCATCTT GATCGAGCTG AATGACAGCC	600
10	CATCATGGA AAAAGAGAGC GAGCTGCTTC GCAAGGTGAA CTCTGCTTCC TCTTCCAGAC	660
	CCCTCTCTGA AATGAGCTCT GAGACATGTC TGAAGGCACT CTTCAGATCC TGAAGGGGCT	720
	CTATGACAC GCAAGCCACA GAAATCAAAA TCAAGCTTGG AACAGGGGAG TGAAGGAGCC	780
15	AAGATGGAG GCAAGAGAGG GTGGCTCTGT TTCCCCAGAG GAAAGCTTAT GACCAATGGG	840
	CAATCGAGCT GAGAGACCTCT GATGTGTGGA AAGGTTCACA GGGATTAAGA GCTTCTCTAC	900
20	TGATGTGAC CCGCTTTCTT GTGTGTGTGT CTGCGCTTGG CTCTCTCTC TACGTTTAAAG	960
	TTTCTGTGAG TATGTTTCTT CATTCTATG CCAAGGTGAG CTGTGTGTTT TCAAGTGTGG	1020
	CCGTCCCGAG CCTGACCTCC AAGCTGATT CTTATCTGGA AATGTACAC TGAATTTCTT	1080
25	GAGTGGCTTT CTGTGGGCC CATGGAGTC AAGCTGGGGG CTGTCTGAG GAGCTGCTTT	1140
	TTTCCAGGGG CCAAGGGGCT GCTTTCTT TGTGTGTATT AACCTTTCA AACATGAGAG	1200
30	GGAATGAGAA GCGCTGTGT CTGACCGGGA GCGAGGTGG CTTGAGAGCT GTGGCGCTCC	1260
	TCTGTCTGT CATGTGAGGT GCGTGGGTGG GAGGAGATTC TGAAGCTCT TGTCTCTTCC	1320
35	CGAGTGGCTC GAGGCTCAC TAACTGGAG GCGAGAGTGA GCTTGCAGG CTGGGAAAGAG	1380
	TCTATCTAAG TGTGTGCTTT GAGTGTGCTT GTGTCTCTC CCGAGAGAA GTTCTCAGAA	1440
	GTTCAGCTTT CCGTTTCTCT TGAATGTGTC TGAATGCCC AACCCAGATC TCTTTCCTTT	1500
40	CTGGGTGTCT TTGCTGGAG GGGGCTGTGT TGTGAGCCTT CCGGTTTCTC AACTGTGCTG	1560
	GCACTTACAG AAGACCTGTGT TTTGTGTAGC GCGCAGCTCT CTTCGTGTTG GCGCTTGTGA	1620
45	AAGCTGAGCC TCCCATTTGT CAGTGTCTGG GTTTGAGCT TATTTGAATG GAAAGATCA	1680
	GTGTTTCTCT GACTCTCAAT TTCTGAGCTC AATTGTCTAC AAGACATGG TGAAGGATTC	1740
	CTGAGAGCAT ATTTCCAGCT GGCACAGAG GGCAGTGTTT GTTCCCACTT AATGTAGTGA	1800
50	CCGATCCAT CCAATGACAG AAGATTATTT TCTGTGCTTG GCAAGAGAG AAGATCTAAG	1860
	GAGAGAGGAC AACTCTACCA GCGAGGTGT TTCCCCAGCA TAAAGCGAGA CAGTTGTGAGC	1920
55	GAAATTCAG AAGCCAGGCA GTTCCGTAAAT GAGCAGGCGCA GTTGTCTAC TGAATGTGTC	1980
	CGTGTGTGTT GGAATGTGAG AAGATCCAG CTGGCAGAGC TGAAGCCAT TGGAGGAGAC	2040
	GTTTGTGGGA GCTCTCAGAA ATCAATGACA AATGCTGGA AAGGACATG CCGAAGTAC	2100
60	TCAAGAGCTC CCAAGATTTC CTGAGGCTTA GCGCAGTGA AAAACAGAA GATCTCATGTT	2160

	TCCAGGGGTC AATCTGTGAG GCAAGAGGA CCGAGATT TTGAGCCAGCT TCAATGTGCA	2220
	GACTGTGAG CTGCGCAGGA CCGAAGATC CAGAGAGAG GCGTGTGTCT GCTCAGCTTG	2280
5	CAGTTCTTGA AAGATGCTG CTGTTTATTC TCCTAAGCTT TTAAGTGGG TGCAGACTTC	2340
	TGTTACAG CTGAGAGCA TTCTTCCAC AACTTTCTCT TCTTGGCCA AAGAGCATC	2400
10	CAGAGGCAAG TAAAGCTGG TTGTTCAAGT ACTGGAGCC GGGGCTGAC TCTTGCAGCT	2460
	GTGCTTAAAG TAAAGATGCT AAAATATCTC CCGTCAATGC TTATATCTCC CTCTCATGTC	2520
	AAAGAGGTA TCTGTGTGTT GTCAAGAGT TCAATGAGT CAGCTGAG CCAAGTGGCC	2580
15	AATGAGCT GATGTGTGAG GTGACATCT GGTACATTGA GGGAGCTGT TTGCTCTCTC	2640
	GACTTAAHA GCAATCATCT TGAAGAGAG GAGAGAGAG GTGTGGGCT AATCTGTGT	2700
20	CGCTGTGAC TTCCCATGCC TTTATGTGAC CCAATGTGT CTGCTGTGCA GAAAGAGAG	2760
	AAGGGCAT TAAAGATGA GGTGATTTAT GTATCTTGA TCCATTTGTC AATTAACATT	2820
25	TTTATTTCTT AAAAATAAAA AAAAATAAAT CAGAGGGGGG CCGGATACCC AATGTGCSK	2880
	AAATGAGAG	2888
30	(2) INFORMATION FOR SEQ ID NO: 62:	
35	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1851 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
40	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 62:	
	CAGTAGTATA ATTATTAATT ATAACTAATT CTGATTTCTT TTCAATTAAT AAGTGTCTTA	60
	GTGCTTAATG AAGTTTGGCC AACTTCACTT GCACTGTGCC CCAACTTTG GAGTGAAGTA	120
45	TGCAGAGTAC AAGTGTGAT GTCAAAAAC TGCATCAAGC TTGTGTGTGA AAGATCTGTT	180
	TTGCGAGATT ATTAATTAATG TATCTTAATG GATGAAAGA CTTTATTA CAACAGATT	240
50	GTCTAGAGCT TCAATGACA TTGCTCTCTA AAGTTTCAAA GTCAAGTGT TTCTGTAGACA	300
	AAGTGTGCA ATTGATGAG GACTTTATT ACAAATTGA TAAAGGATA TCAAGAACTTA	360
55	CAGTCTAATT TTCTCAAGCC AATTGAATTT TCCAAAGCA GTGCTTCTTT AAATGGGAGC	420
	CTGAAGGAC TGCCTTCTCT CCTGTCAATA GAGCTGCCA AACATTTAA CCGAGCTTA	480
	ATTGTCACTC TGCAGAGCT TTTAAGCAAA TGCAGAGCT GTCTGTGACA GAGAAAGTCA	540
60	CTCCAGAGAC AAGAGGCCAA AAGAAAGAAA ACTAAGATG ATGAAGAGAC AACTCCCAT	600

5 AAAAAAGGCGC GTGTATGACG TGAATGAGG CACACTGTAG ACAGCTGCAT CAGTGACATG 660  
AAMACGAAA CAGGGAGGT CCGTAOCCA ACGAGCCTT CTGACATGA GACGAGAC 720  
TCTCTAATTA TTGATCCAG AACTGACAA GATCTCTCTT CCGCTGAAA TACTTCTGTT 780  
AAGAATACC GAATGGAAT TCCATCTTGG TTTTCAGAG ACATGTGAAA TATCAGTGA 840  
CAGCATCCAG AAGACATGC CACATGTT AGATATGAG ATTGTAAAG ATTAAAGAC 900  
CTCCACTGTT CCAAGATTC TACCTTAGCC GAGGAGAAAT CTGATTTGCC TTCTACTTCT 960  
ATCTCTCAG TTCTGTCTGA CTTAGCTGAC TTGAGAGCT GTGATGCCA AGCTTTGCC 1020  
TCCAGAGACC CTAGGTTTC TTATCTCTC AGTTGTGCC ATTTCAGAG ACTCTTTAGT 1080  
CATATGCCAG AACATGACAT TTATGATACC CTGTGTAGGA CCAATTGATC TACAAATCAT 1140  
GTCTGACAA GATATCTGG CAAGGAAAC CAAAGCTGCTT CTTGACATTA GTGTATGAT 1200  
GTCTACTTTT AAGTCCCTCA CCCCAGCC CCATGCTGTT TGTATAGTT TTGCTTTATTT 1260  
GTTTTGTGC TTCAATTTGT CAGTCTCT CTGCTTGAT GCGAAGTAG ATTATAGCC 1320  
TTAATTTCTG CTCAGGAGA ACTCGATG AAAAAAACTT GCATCTTCAG TATACTTCT 1380  
AAGGGCAAT CAGATATGG ATATGTTTAT TGTATTTAAG AGTTCACTTT AGTGGCTTTC 1440  
ATTATATAG CTGTCTGG AGAGACGG TTGCTTAGCC CTGTACAAATG TAAATTAAAC 1500  
TTACAGCAAT TTACTGTGT ATGATATGTT GTCTCTCTG CCAATTTTGT ACCTTATAGA 1560  
GGCAGATGC CTCGATGCC TGTGTTCTT ATTATGAAA TTAGTTTAC TGTATAGG 1620  
ACACACCCA AGAATTTGA TTCTGTAAAG AATCTCTTT AGCTGTGCC TGGCAGTATA 1680  
TAAATGTCG TTTATTTAAC AGATACCTG TCGAGGAAT AAGCAOACT TGAATGAAAA 1740  
ATAATTTT TTATTTTAT GACATGACTG ATTGATGCT ATTCTGTGA CTTAATTTAA 1800  
CTGATTTGA TCACTTAAA AAAAAA AAAA AAAAAA AAAAAAAAA A 1851

## (2) INFORMATION FOR SEQ ID NO: 63:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

55 TCCATATGCTG ATGAGGCTCT TCGTGGCAG GCCAGCTCT TCGCTGTGA TGACCTTGC 60  
ACGAAAAATG CCATCATCT GAAACATTC ATATGATAT CCGTATCAT TGACCTTCT 120  
60 GCACAGGCCA CAGATTCAT TATGAATGA TATAGGATC GTATAGATC ACGAGCCAG 180

5 TTCTGTGATG AGCCTTCAG AAGGACTTA GAGATGCA C TGAATTCGG TAAACCCCTT 240  
CTGTCTCAGG ATGTGGAAAG CTACGATCCA GTTTTGAA CCGTGTCTGAA CCGTGAAATG 300  
CGCGAAGAC GGGGAGAGT GCTGATCACT CTGGGGACC AGGACATAGA CCGTGTCCCA 360  
TGTGTTGTGA TTCTCTGTC CACCGGGAT CCAACTGTG AGTTCCCA CAGATCTCT 420  
TCCCGGTTA CTTTGTAAA CTTCAAGTT ACCGTAGCA GTTTACAA CCGATGTCTA 480  
AATGAATGAC TTAAAGCAG AAGACTGAT GTGACGAGA AACGATCTGA TCTCTTAAA 540  
CTTCAAGGGG AATTCAGCT CGTTTGGT CAGCTGAAA AATCTCTACT ACAAGCTCTG 600  
AACGAGTGA AAGGGGCA TTTGATGAC GACAGATCA TACCACTCT GAGAGACTG 660  
AAGAGAGG CTGAGAGGT CACGAGAAA GTTGAGGAGA CCGACATTT CATCGAGAG 720  
GTGAGAGCG TTCTCCAGCA GTACCTCCG CTCTCCAG CTTGAGCAG CATCTACTTC 780  
ACCATGAGT CCTCAGCA GATACACTTC TTCTAGCAAT ACTGCTCCA GTTTTCTCTG 840  
GACATTTATC ACAGCTCT ATACGAGAC CCGAAGCTGA AGGTTGTCAC CGACACACA 900  
CAGCGCTGTT CCAATTAAC AAGGACCTC TTCCAGTGG GTTTTAACG AGTGGCTGA 960  
GGCATCTGC ATCAGAGCCA CATTAACCTT GCCATGCTGC TGGCAGAT CAATCTGAG 1020  
GGCAGCTGG GCGAGCCAC CTAGCATGCA GAATTCACG ACTTCTTCAG AGGAATAGAG 1080  
ATTCTCTGA GTCTGCTTC CACCCCAAG ATCCAGGGC TCACTGTGGA CGAGGGGAG 1140  
CGCGTGTGA GCTGTAGCTG CTTTCCCGG TTTTAGGACT TGATTTGAAA GTTTTAGGCA 1200  
GACGACAT TTGCACTG GCTGACAGC AGCTCCCGG AGCAGACTGT GCTTACTCTC 1260  
TGAATGTAAG AACACCTGC AACACCAAT GGCAGAGCCA TCCACCGCT GCTCTGATC 1320  
CAGGCTTTCC GCGCGATGG CCGTTTGCC ATGCGCCACA TGTTTGTTTC AACAAACTT 1380  
GGGAGTCTT TCATGTCCAT CATGAGCAG CCGCTCGACC TGACCCAGAT TGTGSCACA 1440  
GAGGTGAAG CCAACACTC TGTCTTAATG TGTCTGTGTC CTGTTATGA TGCAGATGA 1500  
CATCTCAGG ACCTTCAGC CGAGCAGAC ACGCAGATCA CTTCAATTC AATCGCTCT 1560  
CGAGAGGCT TTAAACAGC AGATAGCCA ATAAACAGC CTGTAAAGTC GCGCAGTGG 1620  
GTGATCTGA AGAATGTGA TCTGCCCCA GGGTGGCTGA TGCAGCTGGA GAGAAATTTG 1680  
CATTTCTGC ACCCGATGC CTGCTTCCA CTTTCTCTCA CCAATGAGAT CAAACCCAG 1740  
GTGCTCTGA ATCTGTGTC TCGGGGCGC ATCTTTTGT TCGAGCCACC GCGAGGCTG 1800  
AAGGCCACA TCTGAGGAC GTTCAGCAGC ATTCCGCTCT CACGATATG CAGTCTCTCC 1860  
AACGAGCTG CCGCTTTGA CTTCCTCTG GCTGTGTTT ATCGATCAT CCAAGAAAGC 1920  
TTAGATAG CACCTCTGG GTGTCTAAG AAGTATGAT TTGAGAGTC TGACCTGCG 1980

TCATTTTCCG ATACCGTGA GACTGCTG GATGACCG CAAAGCCAG GCGAAGATC	2040
TCACCGATA AATTCCTCG GTTCGACATA AAGACCTTAA TGGCCAGTC CATTTAGGC	2100
GGGCGCGTG ACAAAGATT TGACCAAGCT GTCTCAACA CCTTCCTGGA GCGCTGTTC	2160
ACAAACAGA GTTTCAGAG TGAATTAA CTGCGATGA AGCTGACCG ACATTAAGC	2220
ATTCAAAAC GAGTGGCAT GAGAGGAGA GAGTTTGTG CATGGGTGG AGTTGCTCC	2280
CGAACCCAG AGCCCTTCCT GGTGGGCTT GCGCAACAC GCGGAGAGG TCCCTCTTAC	2340
CAACAGGCT GTGACATGA TGAATTAAT GCTGAAGAT CAGATTTTG AGATGAGCA	2400
CGACTGGCC TACCGAGAGA CTGAGAGAA GAGAGAGCA GACTCAGAT CCGAGCGCG	2460
CCCTGCTGG ATCGGACAC TCGACACAC GCGCTCCAC TGGCTGACC TCAATCCCA	2520
GACCTGAGC GACCTGAGC GCGCCGTGA GAAATTCAG GATCTTTGT TCAAGCTTT	2580
TGAGAGAGA GTGAGATG GCGAAGCT GCTTCAGAC GTTGGCGAG ACCTTGACA	2640
TTGTCCAGG GTTCCAGAG GAAAGAGAA GAGAGACAC TACTTGACA CCGTTGACA	2700
CGAGTAGTG AAGAGGATC TGGCTGAGG CTGTGTCAG TACAGGTGC CTGCGGAGT	2760
GACCTGATC GAGTGGTGT CCGACTTCG CAGAGAGAT AAGAGCTGC AGAAGATTC	2820
ACTGCGACT GCACTGTGT GCGCAAGGA GCTAAGAAC ATCCAGTGT GCGTGGGTG	2880
CGTGTTCGT CCTAGAGCT AATTCAGTC GACGAGGAG TATGTGCCC AGGCGAGAG	2940
CTGTGCTCG GAGAGCTGT GCTTGAGAT CAAGTTCAC ACTGAGAG GCGCGACCT	3000
TGACCTTCG ACTTCGAG TCAAGGTTT GAACTTCGA GCGGCGAGT GCAACAGAA	3060
CAAGCTGCA GTTTCAGAT CCAATTCAC GCGCTTCCC CTGAGCAGC TGCCTGGCT	3120
CAAGCAGCA AAGACGAGA AAGAGCGAG TTGTGTAAC TTAGCTGCT ACTGAGATT	3180
CAACCTGCA GACTGATC TCAAGTGA CTTCGAAAT GCTACAGAG AGATCTTCG	3240
CAAGCTTAC GAGCGGCTG TCGAGTCTT GTGACAGAG TAACTTTTC TACCTGCCC	3300
TTTCTGAACT ACTGAAAGTT GGTATTAC ATTATTCAT TTTTAAATA TTGGAAGCT	3360
CTGAGCTGT GAAAGAGAG TGGTGTCT GAGTTGAG GAACTGAA GGAATTCAC	3420
GCTTGGAGT GGTGAAAT GAAAGAAC CAGAGAGAT TTGGAAGCC CATGCGTGG	3480
GCTCTTTGA GGAATTAAA CACTAGCAT GAAAAAAA AAAAACTTA GAAACGCA	3540
GG	3542

(2) INFORMATION FOR SEQ ID NO: 64:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 883 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

AGGTGATTT AATGATAGT GTCAATATA GAGAGATA TGTGTTACA TTCTGTCTT	60
CTGATTCAC TCGAGCCG GTTACTAGT GAGAGAAA CAAGATCTT ATTCAAGAA	120
GCTTTACAG CAGCCACAG TCTCATTT CCAATGAAA GATGAGCGC CCGTGAAGTA	180
AGTACAGGC AAGACACT CCGGTTTCT GTTCCGCGCT GATCCAGAT GCGCGATGAA	240
GCGACCGAC GTGTGTTTC TGAGTCCG GTCTGAGA CTAAAGCCCG ACCCGAGAT	300
CGTATGTT GGTGCGAG ACTGAGAG GAAATTCAT CCTTATCCG GTATGTGAG	360
AACAGAGAA ATGTGACAA CCAATGTTT CAGTGAAT CCAAGAGAA AGGAACTCG	420
TGTTTGAAA AATCTGTGA TTATCGTAC CTCTGAAT ATGAGTTTA GATCGAGTTT	480
GACATCTTA TCAATATCC TACTACTCC CAGAAATTC CATTTCTGA GCTGATGGA	540
AAGAGAGAA AATGTACAG GGTGAGAAA ATATCTGTA CCAATCATTT CAACCTTTG	600
TGGGCGCAG AATGTGCCA AATTGACT AGCTCATTC ATGGCTTGG GACTGGGTTC	660
ATGCTGACA GTGAAATCC CTGATCTAT TCGAGAGGC GTATTCACAC ACAAAGAAA	720
ATCGACAAA TGAAGATCA AGCCATGAG GCGAGGCGA GCGACTTTG ATTAGCTAG	780
ATACTAATT CCTGTGATC AACTTAATC CATTTAGTG TTCCCGGAC AATCTCACT	840
CTAGTTGTA CTAGATATG CAGTACATT TTGGGAGAA ACA	883

## (2) INFORMATION FOR SEQ ID NO: 65:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1541 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

GCGAGAGAT GCGCTTACC CTGGGCTCAT CTGGCTTAC AGGACTCTA AAGCTTTCA	60
GATTCCTTG AAGATGACA CCGGACAGG CCTTCACAA GAGAGGAAA ATACATTTT	120
TAAAGCTCG GCTGAGAGA CAGGAGATA CAGAGAGG GTGATGACC CTGACCGAC	180
TAAATGAG GCGAGTGC GCTGTGCTT CAAATAGAC AAGAAATTA ACTGATGTA	240
TATGGGACC AAGAGAGTC CATTAGACC ATGAAAGTA TTATCAAGT GTGACATCC	300

5 TACAGCCGAG GCGCTGATCA TTAAACGAGG ATCCACAGGG TCTGCTGCTT GCGATGAGAA 360  
GATATATGAT CTGATGAGAG AAGATGAGGA AGTGTAGCTG GATCTAGTGC AGCAACATGT 420  
TCCATCCAG GACACCTTCC CTTCTCTGAA CATCATGTGT TCTCCATGCG GCGCAGCCAG 480  
TGTGGCCAT TCCAGTGTGG GCAACTGCGG CCGCGAGGCA GTGTGCCCCA AACTGGAACC 540  
CTGTGAGTGG GAGTATCCCC AGGAGCTTAT ACAGCCCTTC TATAGCTCTC CAGAACTGTG 600  
GATCAGCTCT CTCCCATGA CTGACCTGGA CATCAGTTT CAGTACCGTG GGAAGAGTA 660  
CGGCGAGACC ATGACCTGGA GCAACCTCA GCGCTGCCGA CTCTTCTATG GCGACTGGG 720  
TCCCATGCTT GACCAGAGG AGCTCTTTGG TCCGCTCAGN CTGGAGGAGG TCANATTCCT 780  
AGTCTCTGAG CATATTACCA ATGAGAGCA GAGCTGTTC ACTAGCAGCC TCGTGGAGCT 840  
CATGACAGCA GCACTGATCC TCGAGGTCCAG CGGTGATGCC ATTATATGCC TCAAGCTGTG 900  
CCAGTGCAG GTGTACTGCT CTGGGCCATG TCGCCCATCA CTGTGTTGCT CCAACTGAT 960  
TCAGAGACAA AAGAGGTCA AGCTATTITG TGTGGANCA TTGCTTAGCG ATCTCATTCG 1020  
CCACCGAATA GGACAGATAG AGAAGAGCC ACCGTTTGAG ATCTACTTAT GCTTTGGGGA 1080  
AGAATGGCCA GATGGGAAC CATGGGAAG GAAACTCATC TTGCTTCAGG TCATTCGAT 1140  
AGTGGCTCGG ATGATCTAGG AGATGTTTTC TGTGTATTC ACAGCTGCT TGTATAGTGG 1200  
CAGTGTCCCG CTGCAATCT CACCCGAGA CATCAAGGAT AACTGATTTG CTGAGCTGAA 1260  
GCAAGCTTAC CGATCTCTTC AAACCGAGA GAGCTGGCAG CCGATCAGC CCACCCCGAG 1320  
CATGCAAGTG CCGCTGCCC TCGCTGCCA GTATTTGTGA ATGCAATCTT CTCTCTCTC 1380  
TTTTTTTATA TATGTACAT ATCGATTTT TTATTTGTTA GATTTAACA GCTTTTAAAT 1440  
CTCTGTTTTT TGTGACAGTG TTGAAGTTT GTGATTTCTC AATATGCTT AGATTTAAG 1500  
CTGATTTAAT TTATGGAAA AAAAAAAAAA AAAAAAAAAA A 1541

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 712 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

55 AAAAAAAAAA TGTTHAAGG TCGCTGCCGA GCGCGGACAG AGTGTTTGCT CCGCCTGAG 60  
AAGGCTCTGC TCAGCCCTGA GATGCTCTC CTGCCCCACC GATACTGCCA CTTTAAAG 120

GAGCTGACC GCAJGTGTC GAGAGGATTT GCGCCCCAGA AGATGGGAG TCTGTGCTGT 180  
CCCTTCTGTG TCTGCTGAC CTCACCGAGC CTAGAGGGA GGTGCATTTCA GGTAGATTTT 240  
5 GCTTCTCAT CAAGTTTCTG GCGCTTTGGG CGGAANACAG CCAAGCTTGG GCGTGTGGG 300  
GAGACTCTTC CAGACCGGA ACCCGAGAG GAGACAGAGC CTCGCCATC CTCGACGCC 360  
AGGCCCCGGG CCAAGCTGAT TGGACTGAGA ATTGCGCCAC AACCAATTTG ATGCTGGCTG 420  
GAACGAGAG CCGAAGAGCC TGGCTTTGTC CCAATGTGGG AGCCCTGTCC TCAGCCCTCT 480  
TGTGCTCTTG ACCTGAGTGA ATTCGCCCA GGTGCCAICA GCTGCTGAGC TTCANATTTCT 540  
15 ATATATTCAG AGATTTGGAG AGTATATCAG AGATATTTT GGAAGGAGCT TGGTCTATGC 600  
AATGTCAGTT TCGAATCTTC TTGAAAGTTT AATGTTTTTA TTAGAGATTT TAAAGAAAT 660  
20 AAGGTCTAC ATATCAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 720  
AAAAAAAA AA 732

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 629 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

35 TTAAAGATTT CGGCGCATC CCGCCAGCTA ACATGACTAA AAGAGAGCG GAGATCTGCG 60  
GCTGTGCTCT AGAGATCGAT GCGCTAGAGG AGAAGCTGTC CCAGTGTGCG AGAGACCTGG 120  
40 AGGCGCTGAA CTCGAGACT CACAGCCGGG AGCTGAGCCC AGAGGCCAGG AGGTCCCTGG 180  
AGAGGAGAA AAGACGCTTA ATGACAAAG CTTCCAACTA CAGAGAGGAA CTGAAGTTTC 240  
45 TTGCGAAGA GAACCGAGAG AACATGCTGC TCTGTGTGGC CATCTTTATC CTCCTGAGCC 300  
TGTCTATATG CTACTGAGC ATGTGAGCTT GGCATTTGCC CACAAACGAG ACAGGCTTCC 360  
ACTTGGCCCC TTGCTCAGGA TCAGCAGGC ACTTCAGGCC TCATAGGAG CAGAGTGTG 420  
50 GGTGTGTCCC CTCCCACTCT ATGTTTCAG CATGCTTCC TGGCGGCCA GCGCTTGCCT 480  
CCCTGGCTTG CTGGGGGGTT CCGGCTCTCC AGAAGGAGAT GGTGCTGTGC CTCTCCCTAG 540  
60 CCGAAGGAG AGGCATTA AAGACAAAG CTGAAGAAA AAAAAAAAAA AACTGCTAGG 600  
GGGGGCGCTT ACCCAATGCC CTTTGTGCG 629

60

## (2) INFORMATION FOR SEQ ID NO: 68:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1751 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

10 CTGCTACCCG GCGCGCCGAG GCTGCCAGAC GCGTGAAGCC GCGAGCGCAG CGAAACCCCT 60  
GCTAACCCCG GCGTGCAGGCG CTGAGTCTGC GCGCAGTGGG GCGTCCCGCG GCGTCTCGCG  
15 AGCGTGAATC ATGCGAGGCCC GCGGCTCGGG CAGGGGCGAC GTGTGTCCGC GCATCAGTAC 180  
AGACTTCGAG CTGAGGAGCC TCTCAGCCGG GAGACTCTCT CGGAGAACCA TGCCTCGCGGG 240  
CGAGGAATAT GCGGTGTGAG CGAAGCGTTT CATTCAGCAA GCGAAACTCA TCCCGAGATGA 300  
20 TGTCAATGAT CCGCTGCGCC TTCAATGAGCT GAAAAATCTC ACCGAGTATA GCTGACCTGT 360  
GCATGCTTTT CGAAGGACAC TTCACAGGCG AGAAGCCCTA GATAGAGCTT ATCGAGTCCA 420  
25 CGAGTGAATC AACCTGAATG TCGCCCTTGA GTCATTTAAA CAAAGCCCTA CTGCTCGCTG 480  
GATTCATCCC GCGATGCGCC GAGTCTATTA CATTTAATTC AACCTCCCA AACTCTGGGG 540  
CATTCATGAC CTGAGTGGGG AACCTCTCAT TCGAGCTGAG GATGATTAAC CGAGAGAGGT 600  
30 TATCGAGAGA CTAAAGCGTT ATGAGAACCA AACAAAGCCA GTCCGTGAAT ATTACCGAAA 660  
AAAAAGGCTG CTGGAACATC TCTCCGAGAC AGAAGCGAC AGAGTTGGCG CCTATGATTA 720  
35 TCGTTTCTCA GAACTTAAG TTCACAGGAG AAGCGAGAAA GCTTCAGTTA CTCCAGTAGG 780  
AGAAATGCTG GTACCTATTA ATGATGAAT GCGCAAACT CTTATCTCTT GCATTTCGAA 840  
40 GCTGCTTTTC CTAGAGCTTC TACATGATAT GAACTCTTGG AAAATTAATAT TACTTTTAAT 900  
TCTACTGATT TTATTTTGGG TACTAAGGAT GTGCGAAATG ATCGGATGAC TTAGATGCAAT 960  
GCTTGAATAT CATTAAGCTT GTTGATGACA GTTATCTCA AAACATCGAG CGATGTCTGA 1020  
45 ACGTTTAAAA CATCTGTTAG AGCAAAATTA AAGAGAGATT TCGATGAAT CTACTCTTTT 1080  
GTTCCAGTTAA TAAAGCTTGG ATTAAGTTTC GATATTTTTC TGGAAAGTT AAAAAAGTTT 1140  
50 ACGATGATAT TGGAGAAATC ACGTAACTAG AAATTTGTGC ATGATTTGAT GCGAAAAAG 1200  
ACATTCGAG CATGTGGAA CAGGTGAGA CACTAATATA AATTCCGAAA AGAAGCGAC 1260  
TGCATTTACA GATTATTTGT GAGAGAGAAA TTCAGCTGCG CTTTACACT AAGAAATGTA 1320  
TATGTTAAGC ATTAATGCTG TATTTATTTT GTGCTTAAAG ATACTTCAG TTTCATCAGA 1380  
55 ATTTCATAT TCGATTAAGC ATGATCAAT TACGATTAAG AAAAATTTA CTTTAAAGAG 1440  
ACTGTCTGC TTGAAAAATA CCGTGTGACT GAGGTTTAG ATTGTGTCA AAAATTCAGA 1500  
60

TAAGTCTTT TACAGGAGC AAGTGAAT GAAITTTCA GAAATGTA TTAAGCTTA 1560

TGTTTCAAT TATGATCAG GTTAAGAAAT GTGTTTGAAG ATCTACTGC TGTTTTTCT 1620

TTTGTATCA AATGTGAAT CTGCGCTGAT AAATTAAGAG TTAATTAAGC ATCTCCCCG 1680

CGAATAAAA AAAAAAAGC TCGAGGGGGG GCGCGGTAC GAAITCTCG 1740

NAATGAGAG T 1751

## (2) INFORMATION FOR SEQ ID NO: 69:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 508 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

20 GCGAGAAAT TATGATTTA AATGTTTG AATTTGAAA TATTGAATA TTGTACTAT 60  
25 TTGACCGAC TCAAAATCTC CATGGAAAA TACTGTGCA TACCGAGAT ATTGTGAAA 120  
ATTAATGAT GAGATTTAC AGCTGTGCA GACTGATTA CGAGTTGGCG AATCGAGCA 180  
30 CAGTAAAA TTGAAAACAA AAGATCGAC AACAAAAGC CGAAAGTGG GCGTCAAGAA 240  
GCTGTAGCT GTACTAGCT GTAGAACTCT ATCGAGCT GCGAGGTGTA GTGTGCAAT 300  
35 CGAGGAAAA CTGAGAGAG CCGAGTCTT GAGCTGTGTT TACAGTAG CTCTGTGTA 360  
CGAGAGATG AAGCTTAAG CAGTTTGAAG CTCTGAAGC ATTCAGACAC ATTAAGAGAA 420  
40 ATCGGCAAA GCAATTAAGA AATCTGTTA CTGCTAAGTG TTGCTAGAGC AGAAGAGCT 480  
CTACTCAGC TCGACTTTAA AATTAATA 508

45 (2) INFORMATION FOR SEQ ID NO: 70:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 245 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

55 TACTTAGAC AAGAGAAAT TTCAGAAAT TGTAAATTC TCGAAGAGA ATTTCCTGA 60  
GATTCAGAT TTGCTGTGAT CCTGAGTGA TGATGAGGCG TGTTTTCCC TGTTGCTCT 120  
60 TCTCAGACT CATCTCTCT CTGCTAGCT GTCTGTTGG CAGATGATG TGTCTACTG 180

GCATTTCCTT CACTGATACA AGGAAACTG CAGGTTTAA AAAAAAAAAA 240

NENCG

245

5

(2) INFORMATION FOR SEQ ID NO: 71:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

ATGTTCTCTA TGAGATGCA CTGTGCTTC TCGAGTATT OCTGAGCTT CATAGTACT 60

CCACACGCA CCAGCAATAC AGCTACTAC CTGTGGCTT GGATCTCAGC CAGATGGT 120

GGGAGAGGA GGAGCTGGGC ATGTACCTTA AATCTGTTA CCAGGGAAG ACTCCAGAG 180

TGAGACAG TAGGAGCTTC CTGAGAGGT GGTACAGTG CTCTGTGAT CCACTACTTT 240

TTTTTTTTT TTTTGAGATA GAGTTACAC CTGTGTGCC TGGCTGGAGT GCAATGGTC 300

GATCTGACT CACTGCAAC TCTGTGCTC CCGGTTCAA GTGATCTCC TGGCTGAGC 360

30

T

(2) INFORMATION FOR SEQ ID NO: 72:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 713 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

AGGATCACAC ATAGAGAAC ACTGTATGA CATTTCGGT TCGTCAGAG ACCGAGACA 60

TTGATCACTT TTTGTGTGC GTTATATAT TTTCTGTTAA AAGTTTGTA AAGTTTGTT 120

TTAGCTAGAT GATATTTTAA TAGCTGGAG TCGTTTGGAA CTATTAAGAT GTCACTACT 180

AACACATA CCTATGTTT TGTTTTGTT TGTTTTACAC TCAATATAA TCAGGAGAG 240

TTAGCGAGC ATCTAGCAAT TAGATCTTC TTTTATTG TCTCTTAGG ATATGATGT 300

TCCCATACA CCACAAAC AGCAACAAA ACATTCATA ATATCACTT GATAGACTGT 360

AAGCACCTGC TTAACTTGT GTCCAAATA TTTAGTGTGT ATATATAT ATATATATAC 420

ACACACAC ACATATAT ATCAACAATA AAGCAATA TACATGCAAT TTCACATTT 480

60

GTCTTTCCTT GTTACGNTT TATAGCAGA ACTGTATGAC AGCTTTAGT GATCTTACA 540

TATGTTAAT TCAATTAAT GTAAACAGA TTAACACAA CAAGAACT GTCTATTGA 600

5 GTGAGTCAT GCTTCTAT ATATAACTT GGTTCGGTT ATCCATCA TGCACACTTA 660

TACTGTATC TGATCTTTA TATTAAGAA TACTGTACTT ATAAAAAAA AAA 713

10

(2) INFORMATION FOR SEQ ID NO: 73:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 862 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

GAAAGTCAGA GCTGTCCAT CCTCAGAC CTTTAGATT TCGTCCAAAT TAGAACTGT 60

GGGACTATGT GTTCTGGCA ATCAGAGTC TGGAAATGG CTCGAGGC TCTTGATAGT 120

25 GAGACAGTGG TCATCTTACC AGACATGCAT CTGATTTTAA GCTCAGGCT AATCCACAT 180

CCTCGGCCCT GCTATGAT AAGAAACAAA ACCAAATCT GCTTTTAG TTTAGGAAC 240

30 CTGATAGAA CAGTATTTT CAGCTCTCT GGATAAGCA GTTCTGCAT TTTAAATGG 300

GACTGCAGAA GTGACTGTCT ATAGTTTGA AATACAAAA ATGATATGTT TCATCAGAA 360

AGGAGCCCG TCGCTGGC ACTGAAGAT ACTGAGCTC ATACCTTAA TGAAGAAATG 420

35 TAGGCTCTGT GATGTTAAC TACAATCAG GTTAGGAAG CATATGAC ACCTTTGTCA 480

ACTAAGCTTC ACTAGGAGA CCTGTCTCA TAGAAGATA TCGTTTAAA GTATCAATTT 540

TCCACAGTGG ATGATGAGA AAGTTCAAT TCCACAGAA TCGTATAGT CACATGAC 500

AGCTGACAT ATATACAT ACAGTTTCT GTAAACAGAA GTTCTTCTC TTCAATTCA 560

GGAGTCAGT AGACATAAA TATTCATGT TCACTTAG AACTGATTC ATTTAGAAA 720

45 GGAGATCTGG ATATTTTTC AGGGTAGAAA TGAAGCTAT TTCGGCAT CTGCTCAA 780

AAGTCATAT ATGTACNTA AGTATAAAA AGGTCCTCT TCACCTCTT TGTTCGTAG 840

50 CATTGGCTAC ATACTGTG CC 862

(2) INFORMATION FOR SEQ ID NO: 74:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4602 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

60



(D) TOPOLOGY: Linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

5 GCGAGCGGCGC GKGGGAGACA GCGCCGAGCC GCGTCCCGCC CTTGAGACTA 60  
GGGGGTGAGG GACGAGAG GCGCGATGCC GGGGGAGACG GAGAGGCCA GACCCCCGA 120  
GCGAGAGAG GAGAGAGG GAGAGCGGCC GAGCGCGGCT CGGAGAGAC CGCAGAGAG 180  
GCCCCGAG GCGGTCCCG GCGCGCGTCC AGGAGACACT AGCAGCCCG TCTGAGAGG 240  
AGGTCCGAGC CGAGCCCGG CCGCTGCGGC CGCTGCGGCG GAGGTGTTC CCGCCGAGA 300  
AGGCGAGTA TCCCGCCCG CGAGAGAG GCGCCGCGCC AGCGCTCCCG ACGTCCCGG 360  
GCGAGCGCC AGGCGCGGAA GTCCCGCTCT CGAGTTCAG GAGAGAGAG TCCCGGACTC 420  
CTATCGATAG AAAAGGTAGC AACTGATGAA GATCCGAG GAGAGAGAG GAGAGAGAGC 480  
GATTCGCCC TCCCTCGAGA AGTTTCAGTT GCTGCACTTA GACTAGCCG GCGCTGCGGT 540  
AGTTCAGTA GATCTGTTC TCCGATGCT GATTCAGAGA ACAGCCGAG CTCTCGGTC 600  
AAGAGCGTT GATTCAGCT GATTCGAGG TCGAGACGA ATTCAGACA ACTGATTAAT 660  
GATGTGAG AAGAGATCA GTCTCGAGT GCGATTGTA GTGAGAGAG AGAGAGAGG 720  
GAGAGAGAG TGTTAATAG TGAAGAGAG ATTCATTTCA AGATGATCC AAGAGATAG 780  
ACCTAGAGC CGCAGTTAGA AAGGAGAGC CGAAGGAGC GAGAGAGAT AGGAGAGTA 840  
AAGAGAGAG AAGAGAGAG GAGATTTGAA GTGAGATAG AGGTGAGCT GAGAGAGAG 900  
GAGATGAA TTAGAGAGAG TGAAGAGCT CGAGAGAG GAGAGAGAG AGGAGAGAT 960  
GAGAGAGTC GAGTTTACC CGAAGAGAG AAAAGGCTC GATTCGAGTA TGTCCGTTGT 1020  
GAGAGAGAG GATTCGAGC TGTCCGTTGT GATTCGCTT ATTTCAGACA CGAGTTTAA 1080  
TACGAGCTT TGTGAGAG GAGATATGTA TGTCCGATC GCTCTGTG AGGACTTTC 1140  
AGGCTTACA AGGAGCTTC GCGAGATGCC AAGATGATA GAGTCAAG GATTAATTC 1200  
TGTGATTAAT GTTCCGCGCC CTTGAGAGT TCCGAGATTC TGGAGTACA CGGATTAAT 1260  
GAGCTGCG GAGAGCTTA GATTCGAGA TCTGTGAT TACTTGTCA CGAAGGCT 1320  
CTCTTAATTC GCGAGTAGAG AAGATGATG GAGCTCTT CTAGAGTTT TCTTCAGTA 1380  
TCTGTGAGA AAAATTTAG AGAGAGAG GCGTATGCC AGAGAGAG AAAAGGAGC 1440  
CTAGAGCTT GATTCGAGA GCTGTGCTC GATTCGAG GCGCTGATC ACCAGAGAG 1500  
ATATCTTGG GACTAGACA GATTCCTGA CGAGCTTC AGATGTCAG GGTCTCTC 1560  
TCTCTCTA GCGCTTGA GACTAGCT CTGAGAGTG CTATCTTTA GAGCTGAG 1620  
GATTCGAGA GTGATCTGC AGTGGAGCG AAGCGTGA GCTGATGCT GATTCGAGA 1680

1740 TCTTGTGAG AAGCGGAGC AGTGGAGCA CTGAGGCTT GGTTCAGAG TGAATTAAC 1740  
TGTGTCAG GAGAGAGCT CTGATTAAG ATTCAGCTC TCGCGAGCT TACGTGAGAG 1800  
AAGAGCTG GCGATGAGC AGCTGAGCT TGTATTTA AGTTTAAA GAGAGAGAG 1860  
AAAATTTAA GATTTTAA TGTATGAA TACGTGAG GGTCTCTC TCGATTCG 1920  
ATGAGAGC AGAGTACT AGAGCTTCA CCGTCCGAT CCGTGTTC CTTCTGTTC 1980  
TCCCTTATA AAAATGAT TGTCTTAC AGAAGTGA AGAGAGAG AGAGAGAGCA 2040  
GCTTTAAG TGAAGTTAT TGTATGATC GGTTCGAGC ATTCAGAGAG TCTCTGCA 2100  
TGGAGATTC CCGCTTACA AGCTTACT CTGATGCT CTGATGAGC TTTTACTTT 2160  
TATGATTA TTAATCTTT GTAAATCCT TCTCTCTC TACTGTGCC CTATGCTCT 2220  
GCTCTTAC CCGTGGGCA GACTATGCT GAAATGAT AGATTTAA TCTTGAAAT 2280  
GATTCCTC GATGCTTT TAAAGAGC TGTCTGAG GACTAGAA ATCAGTACA 2340  
TACGTAGT CTTTTGA AGCCTTTA GAGAGAT TTAGAGTAT GGTATCTAT 2400  
ATGCTTTC AGATTTTC AGTTTGA CTTAAGAT TTAGAGGCT GAGAGAGCT 2460  
GTTCAGAT TAAATTAAG ATGATGAT GTTCTGTG ATTTGTTC TTTCTCTTT 2520  
TTTCTTCC TACGTTTC TTAGATTC CTGGGCCC ATCTGCTC CTTCTCTTT 2580  
TGTCTTCC TTTCTTTT GATTCATC GAGTCCCG TTAGAGAG ACAGCTCT 2640  
TATGAGAG ACAGATCTA TGAATCTTA ATTTTACT CAAGAGAG TCGTCTTT 2700  
CGCTTTAT AAGCTTTC AAGAGAGAA ACTTCAGAG CGAAGTTTA TATCTTATA 2760  
TATTAATTC TTTGCTTC ACTTAATTT CTGGGATCA GATTTCTT GGTATGAG 2820  
TGTGTGAG GATTAATTC CTTAATCTG TTAGCTGAG AGAAGGTA GTTCTTTGT 2880  
AGAGGCTAT ACCGAGAA GCGCGATTC GTATTTGCT TCAATCTG GTAAGAGTA 2940  
ATTCCTATC TGTGTCTC AGAGTGA ATCGAGGCT TACCTTATC ATTTTATTT 3000  
GTGAGATG CAGTATGCT CTTGAGAG ATGAGTTT GCTGAGAAA AAAAGAGAA 3060  
GAGTTTGT TTAGAGAA TGGCTTACA ATTTTACA AGATTTCT TTTCCGAAA 3120  
AGAGAGTA GAAATGTA TTTGAGAG AGGCTTACT TTAGAGACT AGTTCAGCA 3180  
TTTGAATC GAGGAGAG AGATGAGC ACTTCAGATC AGAGTCTA AATGCGTAT 3240  
TGTCTTTT GAGATGTC GAGTTTCC ATTTCTAT AATTCAGCA TGAAGAGAG 3300  
AATTAAGTA GAGGATAT AGATGATC TTTCTGAA GATGTTCG AGTGCATTT 3360  
GATTAATTA CTAGATTGA AATTAAGTA TTAGATGAT GATTCAGAA AAAAGTTTC 3420  
CTCTCTAG GTCCCTTTA GATTTTGA CTAGTACA TGTATTTCC AGCTTATCT 3480

TTTCTTCAC ACTTATCTT TGTCTCGTA AATTTCATT GAGTGGTTA GTCAATCAGAT 3540  
ATTATTAGCA CCTACAGAAA AGCAACTGCG ATTTTAAAAA ATCTTTCTGA GATGGGAGAA 3600  
5 AATGTATCTT CCTTCTCAT ACCGCTGCG CAACAAA AAAA ACACTACTT AGTTCTACTA 3660  
ATTAGAAACT TCGTGTACTT TTTCTTTCT TTTAGGGGTC AAGGACCTC TTTATTAGCTA 3720  
3780 CCAATTGCGT ACATTAATTT ATTGACAGAG TTTGCAATAC TAAATATTT TTTATTAGCT  
3840 TTTATATTTT CCTTTTGATA AAGGATGCT GCATAGTAGA GTGGTGTA TTAATACTATC  
3900 TCAGCCGTTT CCTGCTTTTC CTTCTGCTC CATATGCTC ATGTGCTTC CAGGGAGCTC  
15 TTTTAATCTT AAGTCTTAC ATTTCATGCT CTTCATGAAA TTCTGTACC TTTTAAATTA 3960  
CTCTTCCAC TCATATATTC CATCTCGAT TCGTGTCTT AATTTCTGAA ACTGTAGTG 4020  
4080 AGATACAGCT ATTTAATATT TCTGGAGAT GTGCATGCTT CTCTTTGTG GTTCCCGAG  
20 GTGTGTTTGC GTACTGAGA CTCCTGATA TCGTTCAGAG ATTTAGGCA AACACTGGCC 4140  
ATGGCCGTGG GAGTACTGGG AGTAATATAA AATATGCG GTATAGACTA GCATCCAGAT 4200  
25 AGAGCACTTG AACCTCTTT GTACCTGTTT GGGGAAAAG TATATAGAT GTACTAGCAA 4260  
4320 TCTAATAG ATTTATATAG TCTGTTGTT TGAATAGCA TTTTCTTC CTTTGTGTT  
4380 TTTCCGCTT TCCATGTA TCAAGAAAT TGAACAATG TATGATCA ATTTAAATA  
30 TTTTATTCTT TAAAGGCTT TTTTGGCTT TGTATGTGC AAGACCTTC TCTTTTCAG 4440  
4500 GGAGAGAG GTAGTACTT GATATAGT TGAAGAAGTT ATGTAAAG ATTTATAT  
35 AAAAGGATA CTTTGTCTT CAATCTTTG TTTTCTCTTA TTCTAGGTA GCATATTA 4560  
4602 AATAAATAT GTAAAGAGA AAATAAAG TGTCTTCAT G

## (2) INFORMATION FOR SEQ ID NO: 75:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1255 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CGCCGCCGG GCGCGCGGT TTCTTACA AATAAAGA ACCGCACTG CCCAGCGAG 60  
CGTTGCACT TTAAGATGG TCGCTGGG GAGCTAGCC TCATCTGAT GATCTGCA 120  
55 AGGCGCACT TTTATTTTA TTTATTTT ATTTTCTT TAGCACTCT TGGGGCTTC 180  
ACTCTCAGAG CAGTTTTHA AGGAGACA GAGCCGAGC CTGCTGAT TCTATGCTT 240  
60 GGTGTGACT ATAAGAGTA TTGCTAAT TCAATTTTA TCTCTTAA CAACTGTG 300

GCAAAAGAT ATTGACCGT TTCAAAATT CAGATCTGC CTCTGCGAT AATATTTC 360  
CAGGATGAG TACTCTGCT CACACTCTG AAGTCCAGA CAGAGGTTT TACACATTC 420  
5 TTAGCACTGA ACTGCTGCT GATCTAGAT GATCTGTCC CCTCTGAT GAACTGCTC 480  
TGATGATA GGCCTCAGC AGGCTACTTT GAGGGGACA ATCAGATGCA AAGGCTCTTG 540  
10 GGTGTTTAT TAAATACTA GTGTCACTTT CTGAGTACC GCGCTTCA AGGCTGATC 600  
CAGGCGTGG TCGTTGTAG AGCGAGTGC TTGCTCAGAG CCACATTTCC ATTTCATCA 660  
720 TACTGCTTT CACTGCTATA GTCACTCTTT TGAATCTGG GAACCAAAAT GTGATGATA  
780 TATAGACTTT ACTATAGCC AAGTTCATC CCACGCTTA GTCTTGGAA TGTATATAT  
840 TGAATATCT AGAAATGCA TTCTATCAAT TACAGATTC AATATITGCA AAGGATGTC  
900 TGTCTTCTC CCGAGCTGC CCGTTGCCC TTAATGAAA ACCACACGG TGCATCTCT  
960 TGTGATGCA GCGCTATGCA CCGCAGGCA CCGTGTATG CACTCCGCC TTGTCTTAC  
1020 ACAGCTGG GGTGTTAGS CAGTCTGCT TTTTCACTT AATAATACAG CTTCAGAGA  
1080 TTTTGTATC ACATTAATA TCCACTGCG TCTTTTGTAT GGCACATAA TAACTACTGC  
1140 ATATATGCA TAGCGCTAT TGTATTAAC TACTTCCCTA ATGATGACT TTTAGTGTG  
1200 TTCTTTTCT TTTCTTTT GCTAGTCAA AGATGCTAT AATAATGTC CTTATCAAA  
1255 AAAAAAAAA AAAAAAAAA AAAAAAAAA NCGGGGGGGG CCGCGGAGAC NCAT

## (2) INFORMATION FOR SEQ ID NO: 76:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 475 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

GGCAGAGAG AAATGTTGA TTCTTTTC TATTTTAAG GATCTTCTT CTGTGTGATG 60  
120 TTGAAACTT ACCTTAGTGA AGATGTTT CAGATGCTG TTGTCTTTA CTGCATAT  
180 CAGACTATG CATCTATCA AGTGTATAT CTGTGGGATA GTTTTATGCA GGTCAACAC  
240 CAACTAG ATGTAAAGAG AATGATGAA ACTTGGACC TCAGAAAGG ATTTCCTTTA  
300 GTGACTTTC AAAGAAAGG AAGGAACTT TTTATCAAC AAGAGATTT CTTTTAAAT  
360 ATGAAGCTTG AATTCAGCC TTGAGTACA AGGTACTGCG CCTCTTCTT TTCATGCCAT  
420 CTCCTTTGCA CTCACGCTG GAATATTTT GAGTCTTTT AATATCATAA GTTCTGTGA

AACTTAACA GATTACCTT TCTTAAGAT ACTTAACCTT GCTACCAAT TAAAA 475

5 (2) INFORMATION FOR SEQ ID NO: 77:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 465 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

15 TTCTCTTGC TCTTGACTG CACCCACTG GCGGCTGAC CTGACTCCC CTAGTACT  
CAGCGGTGCT GCGATGGCGT GCGGCGCGCG GAACTGCGG TCGGGGCTGG GCGCGTGG  
20 GCTGTGGCGT TCGTGGCTT GCGCTGTGCG GTGCGCGGCG CGGCGGCGCG GCTGTGCG  
TGTGTCTCGG CGGTGTAAAT CATTCAGTAC GTGAACTCC AGACCAACT GACGCTGTGG  
AGCGTCTCGG AAGTGGCGCG CTTCGCGGAC AGCTGCGCCA AAGAGGGGCG GCGTGGCTGG  
25 GTGGCGCTTC CGTGGCGCGC GCGCGGAGAG CTGAAAGGCT KCGCGCGCGA CACCGCTTTC  
TTGCGCGCGC AGCGCGGCGG CGAGGGGCGC GCGCGCTGGG TCGCGCTGCT GGTGTGGCGG  
30 GCTGCACTT TGAAGAGAA AGTGTGTGGT GCGCGCGAGG AAGAA 465

35 (2) INFORMATION FOR SEQ ID NO: 78:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1907 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

45 AATGCGAGCG CACTACAGA TTCTTAAGA ATTCTGAG GTGCGAGAA GAAATAGAG  
AAGCAACTG GAAAGATCC AAGAGAGCT AAGTGTGTTG GAGAGAGTA TTAAAGAGT  
50 GAGAGAAATG AGTGAATTAT ACTGTGCTGT CAGTGAGAT AACAAGTGC CTGAAATTGA  
AGCTCTCTCT CAGTACAGA GTAGTATTAT TGAATTCACA GAAATAGCGC AACTGCGAG  
TTTCAGTGC AGTGTGAGA CAAGAGACA GCGTGTGTTT AATAGCACT TACAGTCAG  
55 ACGAAGAGCA CTGACTGCTC ATTGTAGAGA CTTCAGGAG TGTACTTTT CTACAGAGAT  
GCTGTGATTC TGAATAGACA GTGAACTGC AAGCAAGTTG GATGAATTC AGAATGCTT  
60 GTCCAGGTTT ACTGATATA ATTCAATAG AACTTGAAC AATATTCAT ATGCTAGTA 480

TTCTTAATAT GGTTCAGATA TACGTCTGAG TATTAAATT GACCGGAAAT GTGACTATT 540  
TGCATTGCT GAGATTACAA AAGATTTAA AGCTTAAGA TATGAGACTG TCAATCAGGA 600  
TGCATGCAAT ATTCAATTAC CTGAGAGTAA AATGACTGC AATTGAGAAA TCACTGTAT 660  
5 CAGTGAAGT AATTACATA AAGACTGTTT AGCTGACAT GATTAGAGAG GACTGTATAT 720

5 (2) INFORMATION FOR SEQ ID NO: 79:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1168 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

15 TAAATGTGCG TGTGTTAAAT TCAAGCGCTC TTGCAATGAC GATTGGCTT TCGGCTGCG  
AATGCACTG GTGCACTACT ATGATCTTGG TAACTATAAA CAGCAATGCA TGTATTTCAA  
AGAGACCGT AAGAGACTCT CTATGCGAAA GTTGTGAGT GTGAGAGAAA TTGTCTCTGC  
20 CTGACAGAGC AGTGAAGCTAA AACTGTGAAA TGTAGAGAAA CGATAGTGCCT TACGTCTCTT  
CAGCGGTGAT ATGATATGAA AAACTTGTGT AGCGTGCCT TCGAATGAGG ATTATATGAC  
TTGTGAAAT GAAATATGAT CTCTGACT GTACTTAAA GACTTTTCA AACTTGTGCT  
25 AACTTAAAG TTGATAGAG TCAAAAGTGT TCTGACAAA GAGCGAAGAG AAGATGATAC  
AAATGAAATT GTTATGCTG TGTCTGAG GGCATACAGA GATGGGAGCT CGAATGTGCT  
30 GATTCCTCT AACGTGAGG GTACATTTAA GGTCTGAGAA TTGTATGATA GGTATGACTC  
AAGTCAAAAT GTACTGATC CTGCTGAAT AACTGTGAG CTGACATGCA GAGAGAGAAC 1500  
AAGAAATGTC AATGATATGTC TGTCCCAAAA GTCAATGATG GTTGTGATTT TGTGTAAT 1560  
35 AATTTTCTCT TTTTCTCTTT TCCCTCTTTT ATGACTCTTT GACAATGGG AATACCGAGC 1620  
CAACTTCCA CGATCAATGT AACTGCGATG ACATGTGCTC TGTGTGTGCT GTTATCTAAT 1680  
40 TTTTGTGATA GCGAAGAGAA TTCTTTGATA TAAATATAAA TAAAGAGACA ATAAAGTTT 1740  
ATTGAGCCGC AGTGAAGCTT GAAAGTTT TTTCATATGC TCGAAGAGT AACTCTTTT 1800  
45 AATGATGAGC AATATGTGAC TAAATGTGTA CAGTGATATA TTGTAAATTT TGTATTTTCC 1860  
CAACTCTTT GCGAATTTAA CATATGATA TAAACAAAT ATAAAGT 1907

50 (2) INFORMATION FOR SEQ ID NO: 79:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1168 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

5 OCTGGGGTGT CCCCCKSGCC ACCATGTCTA TCGCTTACTT GHTGAGGCAC ACTCGGHTGA 60  
CCCATGCTG ATCGTTATTA ATTTGTCTAA GGCNAAGCAC CATTATCTC CCCAAACCTT 120  
AAGTTCATG GCGAGTGTCT AGAGTTCGAG GAAGACCTTA ACACGCGTGT GACACCGAGA 180  
ATCTTACAC CAAGCTGAT GGGCGTGGAG ACCGTTGTGT GACATGCTC TCGATCGAAA 240  
10 GGAATGCTGC TCTCCATTAG GAGCAATCA GGAAGGAGA TGGATCTCG TTTTITTTCT 300  
TTCTTTTITTT TTTTGTAGTT GGGAGTAAGT TTGTGATGG AACAAACTT GTTTAAACAC 360  
TTTATTTTAA ACAGTGTAA GAAGACTATA ACTTTTGATG CCAITGAGAT TCAGCTCCCA 420  
15 CAAGCTGACA AATTAGGAG GTTAAGGAG TAATTTTITTT AAGCCACAA TAAATATATA 480  
ATACAACTTG TTTCTGCCCC TTTTCTCTTT AAGCTATTTG TAGAGTTTAT GACTAAATAG 540  
20 TCTGTGAGG TTTATAGACC GAAGTACTA CACACTTTAA ACCAATTAAA AAGAACGAAA 600  
AGTAAATAGA AAGACATTTG AATCACGAG CCGTGGGATC AAGCTGGGCT GTCCACACAG 660  
AAGACAAAA CCCAACAAA CCAAGCCCTG TTCTGCTCAG TCGTGCAGAG AAGAGATCAG 720  
25 GGCAGCTTAA GTGGCTTAG RATCTTTGAG GCATTTCTTTA AGGAGAAAA GGAATCCTTT 780  
GATTTGTGT GTTTCAGCT CTGGATTTT TTTTITTTT CTCTCTGGG TTTAGAGAT 840  
30 TTTTITTTGAA ATAGTGAGA ACTGACCATI ATATGCTTTC ACTGGCTTCT TGTGCATATA 900  
TATGATGTTT TANGTGTGCA AACAGTTAG AGCTGGCAGC TGAATGATAG ACAATAGTG 960  
35 CAATTTGCC AGCTTGAGAG TAGAAGAGAA TTCAACATTA TATCAATATC TTTCTTCCC 1020  
ACCTTTTCCC TTTTITTTT TTTTITTTG TTTGATCTG GTTACAGTGC CATAAACCTT 1080  
GTTACATATG TATATCAGAA TGTAAAGAAA AAAAATTTAT TTAATAATAT TTTTGCAAA 1140  
40 AAAAAAAAA AAAAACTGCA GGGGGGCC 1168

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1285 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

55 AAAAAATCAC ATCTTAACAA AGAGTCTGT CTAAGACAGT ACATCTCTG TTGAATCTGC 60  
ATCTTTCCAC AGGACTTCT GTTTTATAGG ATGAGACTAT TCTCTCTTC ATCAAGGAAA 120  
GAGAAATGTT CAGGGTTATA GGAATGGCAC ACTTATTAGT TCTGCTCTC TGAAGGTTTC 180

CTCAGGACA GTTTGTCAG AGCTGCATTT CTATGTCCAT GGTCTAATCC TTGAGTATCT 240  
CTTTTTCOC TTTCTGTCT CAGGAATCAG CTGAGAAATC AATCGATGTT CATGCCCTTA 300  
5 GCGCTTACT GTGATTTGTT GGTTCACCTT TCATTTGCTT TAGTTCTAGA ATCACTGTTT 360  
GACTCTCAG ACTTCACTTA ACTTTGAAA CTCTCTTTTG GAGGCTTCTC AATTCCOCCCT 420  
AATCTGTGC TCGCTGAGCC CTAGAAATTT CCCACGACG AATTAATCCA GGTAGATCCT 480  
10 AAGTGTCTG ATCTAGTTGA TATTTAACA ATATCTAGTT GATATTTCTC ATTCAATGCG 540  
ATCCAGAAC CAGTATCTCT NAAAAACAAC CTCTCATACC TTGTGACCT AATTTTGTGT 600  
15 GCGTGTGTGT GTGGGGCAT ATGTATATAG AGAGGCAT CTTTTITTACT TTTGTAAAG 660  
CTTATGCTC TTTGTATCT ATATCTGTGA AAGTTTAAAT GATCTGCCAT AATGCTTGG 720  
780 GGCCTTGTCT CTCTGTGTA AATGTAATA GAGAAACAC CTATATATAG AGTCAATCTA  
20 GTTGTTTTAA TTGCATGA AGAAATTTT CAGATACAA CACTAACAA CTCTCCCTTG 840  
ACTAGGGGA CAAGAAAG CAAGCTGC CATAAAAAC AATTACCTGG TGAAGATTC 900  
25 CATTAACAGA ATTAGTAGT ATATGAGA CAGCATCAT AAACAGTTAT GTTGTCTCC 960  
TTCCAAAAA CATGTACTGA CTTCCTGTTG AGTAATGCGA AGTTGTTTT TTTATATATA 1020  
1080 AACTTGCCT TCAATCATG TTTCAAGTG GTGTGTGGG CCMAATATY GAATGATGG  
30 AACTGACTGA TAAAGCTGA CAATAAGCA GTGTGCTTA CAAGCAAC AGTAATGTTG 1140  
ACATCTTAA TTCAAAATG CTAAATTCAT TATAATGTGT TTTGTAAAA TACACTTTGA 1200  
1260 AACTATTTT CTGATTTCCA AGAGCTGAGA TCTTAGATTT TATGTAGTAT TAGTGAAAA 1285  
AATACGAAA TAAATACAT TGAAG

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

TTCTCAGCCC CAATTTCTAC GGCACGCGA AGCGGAGGT CTTCTTTCT TCGCTNAGCC 60  
AGCATGGCT COTGTGCCA AGAAGCTCT GAAGCGGTG GCAGCTCCA AGCATTTGAT 120  
55 GCTGATATA TTGACCGGTG TGTTTGCTC TGTTCATCC ACCGGTCCC ACAAGTTGAG 180  
AGAGTGTCT CCCCTATCA TTTTCTGAG GAACAGCTT AAGTATGCC TCACAGGAGA 240  
60 TGAAGTAAG AAGATTGCA TCCAGCGTT CATTAATATC GATGCAAGG TCCGAACTGA 300

360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1290

TTAATCAATAC CACTCTGAT TCAATGATAT CATACAGATT GACAGACAG GAGAGAAATT  
CCCTCTGATC TATGACACCA AAGGTCTGCTT TACTGTACAT GATATTACAC CTGAGAGAGC  
CAAGTACAG TTGTCCAGAG TGAAGAAAT CTTTGTGAGG AGAAAGGAA TCCCTCATCT  
GATGATCAT GATCCCGCA CCATTCGCTA CCCGATCC CTCAACAGG TGAATGATAC  
CATTCAGATT GATTAGAGA CTGCAAGAT TACTGATTTC ATCAAGTTCC ATTACCCAG  
CCAGCTGCTC TGTACACTC AGAGCTCTCG GAGATCTCTG CCCAGGCGAG GACTAGAGCA  
AGCTTCAGAG GACTCTAGG AGCTGCTCT TCTACAGAG ATGACTACAC TGTATTCTTG  
GCACTACTG CTTTCTCTCT GTACACCCA CTTTGAGGAG CCATTGAAA AGGTGACCTC  
TGTGGGAAT TCTAGACCA CAGGCAACA GCTAGAAATC CTGGGCTGCC TGGCCGCSGG  
GAGACAGAC CTGCGCTGCA CCGAGAGAA GCGAGCTCT ACTCCAGAC TGAACCTCG  
GGGAGCTTG CTGTCCCGC CCCCGAGAG CTCCGAGAC CCCAGACAG CGGCTCTGTC  
CGCCCCCAG ACCCCCGAGA TCCCTCGAC CCAGGCGCGG GTGCTGTGC ACCGAGAGAA  
GAGCTGCC AGCTACACT GAGATCTTT CGGCTCTGCG TTCCAGAGC GAGAGCGGC  
ACGAGAAC GACGCGAGAA GCGCTGCGCG GCGCTGAGAG CGAGGTGCG GGGCATGTAA  
CTTCAGACC CAGAGAGTTC AGAGCATCGG GCGCGCGCGC GCGCGCGCG GAGTGAAGGC  
TGAAGAGGG GCGCTGAGG CTTCAGACC GAGGCATTAA AAGAAATTT GGTATACCA  
AAAAAAAA AAAAAAAAAA TCGCGCGCGG

40 (2) INFORMATION FOR SEQ ID NO: 82:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 684 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

60  
120  
180  
240  
300  
360

TTTATGTAT TCTATACTA TGAATCTT ATTATATCT TTTTGGACT TGTATAGTG  
TCTTATAGG TTTTATAGT CAGCTGAG TTTCAGATAT TGACTATCC CCGTAAAGAT  
GAGTTTATT ATAGACTTA ATGATGAAA AATCTTACAT CTTTGAGTAC TCTCTTCAAT  
TTGTCACTG TTTCATGTGA TTCTAACCA TGGTATTTTA ATTCTGATT TTTTCTTTC  
TGTAAAGAA CATCTTTGA AAAATTAATT GAGAGAAAT TGAATCTTA TGAAGAGGC  
ATTACTGAC AGAGAAAT TTTTGTATT ACCAGTCCC TAAAGATGCT AACAGCTGAG

420  
480  
540  
600  
660  
684

GAGCAGTAC AACACAGAT GATGAGAAA TCCCGAAT CCGTTTTC TTGSCCAT  
CTTTCTCC AACCTGAG GAAATACAT YCATATCTTA ACTGAGCTT AAAAGTGGT  
TTATCAGAC CGACTTGA GGGCTGAG CTTHAGTAC TGTACCCCA TCAATACGA  
GCTTACGTA TTGATCTCT TTTATCTT CAGATTTC TTTAAATC TTGCTTTT  
TTTCTTCC GAAAGATCC CCAGACTTA CCATTCCCA CTTCCTTG AATTTTTG  
GCTCATTT TGAATTTTC AAGA

15 (2) INFORMATION FOR SEQ ID NO: 83:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2024 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080

CTGAGAAAT TCGACAGAC TCGCTGAG GCTTACAT TCGCGCGCT GCGTGCCT  
TCTGGAAT GAGTCTGA GCTTCTTG TTCTGTGCT GCGCGCTTG CCGCTTCTC  
GCGCTGCG GCTGAGAC TGGGAGAGG GTTGAAGGG GCTGTGAT CCGCCGTTA  
AGTTCGCTC GCGCGCGCA TGTGCGCTG CAGGTGAG GCGTATGT CCGAGCTGAG  
CGCGGAGC GAGGGGATG AGAGAGAA GTGCTCTAT GCGATGAAA ATGAGTTGA  
AAGCCAGAA GAGGAAATG CAGTCTTA TCTCTCAT GAAATTGAG ATGAACTGC  
TGAATATCT GTACAGAAC CGAAATGAC TGAAGCGAA GATGATGAG ATAGTACAG  
GATGATAT GAGATGATG TGCATGAC TATGAGAC ATTAAAGCG GAGCAGCA  
GTATGGAAT TATGTACAG CACTGTAA TCTTACATC AAGACAGGG GAGAGTTTA  
TGAATACCA GGAAGAAA TGAAGAGT AAGCTTAT GAGCTTGA GATTTATG  
AGTTCAGT TTAGAGTATG ATTGATATC TTTTGAAAT AAACATGCG GTAACTGCG  
TCTGATCT TGTGATAT TTAATTAG GTTAAAGAA GATCTGGA AAGCTTATG  
TGAAGAAA AAGAGATAC GATGAGAT TGAATTTA CAGTATAC CTATGAGAA  
TAAATTAG GTACAGAG GAGAGCTG AATCTGAG AAAAGAACT CCGTTCATC  
TGAAGACT GAGTTACT CTCTCTTC TTGTCCAG ACTGCGTTC GACCGACAG  
GAGTTACT GCGGAGATG ATGTATGCG TGAAGTAA ACTATGAGC GAGTGAAGG  
CAGGAGCG GAAATGAA AAGACAGAT AAGATCTCT TCTGAAGAT CTCTACTGA  
ATGAGACAT AATTGACA AACCATCC GTTTTCTCT CAGAGACT CTCCACTCA

5 CTTTCCAGCT CCTGCATTTTC TTCCACCTCC TCCGACTGTC AGCACTGCTC CAGCTCTGAT 1140  
TTCCACACCG GGTTTTTCCTC CTCACACGAG CCGTCCAGCT CCATCTCTTA TACCAACAT 1200  
AGAAAGTGA CXTTCTCTG GTTATGATAG TCGTTCTGCA CTCGCATTTTC CATATGCGAA 1260  
TGTTCCTTTT CCGCATCTTC CTGCTTCTCC TCCTTGCTGG CCTAGTCTTG TCGACACAG 1320  
CAAGCACTGG GACTATATATG CCGAAGAGA GAAAGACGA GATAGAGAGA GAGACAGAGA 1380  
CAGAGAGGA GACCTGTATC GGGAGAGAGA AAGAGAACGC ACCAGAGAGA GAGAGAGCGA 1440  
GGCTGATCAC AGTCTTACAC CAGTGTTTT CAACAGCGAT GAAGAACGAT ACAGATACAG 1500  
CGAATATGCA GAAGAGGTTT ATGAGCTCA CAGAGCAAT CAGAGAAAG AAGAACGACA 1560  
TAGAGAAAGA CGACAGAGGG AGAAGAGGA AACCGAGAT AAGTCTTCTC GAAGTATAG 1620  
TAGACCTGCG CATGAAGCTG AAGAGAGGA TAGTCAAGG AGACACAAAC ACAAAAATC 1680  
TAAAGAAAC AAGAAAGGAA AAGAAACGCG CAGTGAAGCT GCGCTCTGAC AGGAGAGCAC 1740  
CGAAGCTACA CCGCAGAAAT AGGCATGTTT TTGGCTTTT GTGTATATTA GTACAGAG 1800  
TAGATACTAT AATCTCTT ATTTTCTGG ATATGTTTA AGAATTTTAC CTTAATCTT 1860  
GTTCTGTTTG TTATATGAA AAGTAACTT TTTTTCGAA ATAAAGAGT GAATTTTCA 1920  
TGTTAAGTTA AAATCTTTG TCTGTACTA TTTCAAAAT AAAGAGAG CAGTACTTT 1980  
ATATCCAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGC GCCC 2024

## (2) INFORMATION FOR SEQ ID NO: 84:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 911 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

5 CCGGCGAATA GCGCGAAGGG GATCTGAGCT GCGAGATGA ATGTGGGGT GGCACACAGC 60  
GAAGTAAGCC CGACACCGG AGTGAATGAT AGCGGAGGCA TCTGGCTGGC CTACATATC 120  
TTGGTAGGAT TCGTGCATAT GGTCTTACT AGCATCCCTT TCTTCAGCAT TCTGTGTCTC 180  
TGGACCTTGA CGAACTCAT CCAATACCTG GCTAGCTATG TCTTCTTCA TACGGTGAAA 240  
GGGACACCTT TTGAGACTCC TGACAGAGA AAGGCTGGGC TACTGACACA CTGGGAGCAA 300  
ATGCACTATG GCGTCCAGTT TACTCTTCC CCGAAGTTCC TCAGCATCTC TCTATTTCTG 360  
CTCTATCTCC TGGCAGCTT CTATACGAG TATGATGCTG CGCACTTCTT CATCAACACA 420

5 GCGTCATTGC TAAGTGTACT GCTGCGGAG TTGCCCCAGT TCCATGGGCT TCGTGTCTTT 480  
GGCATCAACA AATACTGAGG GATGGGTTTT GGGAGAGCTC CATGGGCAAT GGGAAAGGAC 540  
TGAACAGAG GACTATATTA CATCTCTTC TTATCTCCA TACTGTCTTC TAGACCTTTA 600  
AAGCTGAGA ACTATACAC CTTTCCAGA CTCACAGAA GAGAAGAT TCGCAAAATGG 660  
GGCTCTGGGG CCCAGTCTTG CTAGTGCAA GTTTCTTTGA ATCAGAGAGG CAGGTGAGST 720  
AAGGCCAAA TCATCTCTCT CATTAGCAGG AAGCAATTTG GGCAGCTCTT TTGCTGATTA 780  
CATCTTTCCA TATCTTTTAC ACTTACACC TTCCAGTCTT GTTTTGTCTT GTATTTTCTT 840  
TACATAAAT TTTTTCAGCT ATAGCTGCGG TTTAATCAGG ATGGGTAGAG AGCTGTCTTC 900  
ATAGGCTGG GGGTGGGAG ATGGAATACT G 931

## (2) INFORMATION FOR SEQ ID NO: 85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 825 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

5 CCGGCGCGCG GGGGTCTTCA GGGTACCGGG CTGGTTACAG CAGCTCTTACC CTTCACGAGC 60  
CAAACTGGC AGCGCAGAG GACCAAGAGA AAGATCCGA GCGCGAAGG CTGAGCGGCA 120  
CGACCTGCTT CCGGAAGCTG ATTGCTCTCG GTGAGGGCG GAGTGGCTG GAGCGGGCGC 180  
GGCGACCAT CCGGGCTGAG AGCACTTTCG TGGACACGA GCGCTTCTCA CGGCGCGGCA 240  
ACCTGGAGA GCTGTGCCAG CGCTCTGAC GCACCTGGA GTACTACAG AGCACTATG 300  
TGTTCGTGTT CTGCGGCTC ATCTGTACT GTGTGTGAC GTCCCTATG TTGCTGGTGG 360  
CTTGGCTGT CTTTTTCGGC GCTCTGTACA TTCTCTTACT GGGCACTTG GAGTCAAGC 420  
TTGTCTCTT TGGCGAGAG GTAGGCCAG CGATCAGTA TCTCTGGCT GAGGGCAATCT 480  
CTTTCCTTT CTTCGTGCTG GCTGTGGCG GCTGGGCGCT CTTCGTGGTG CTGGGAGCCA 540  
CCCTGTGCT CATCGCTCC CAGCTGCTT TCCACAGAT TGAAGCTGTG CACGGGGAGG 600  
AGCTCAGAT GGAACCGTG TGAAGTGTCT TGTGGACCT GCGGCGCTCC CGGGCGAGCT 660  
GCGGCGCGG TCGGCATGCC TGTCTGTGAG GGTCTGTGCTG CTGGGGCCA CAGCGCGCTC 720  
CCATCAGAG CCGGGGAGG GATCCGCGCT TTGAATAAA AGCTGTTATG GGTGTGATTC 780  
AGAAAAAA AAAAAAAGG GGGGCGGCTC TAGGGGTCAA AGTTA 825

## (2) INFORMATION FOR SEQ ID NO: 86:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1238 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

CACTGAAAG GATGAAATG GACTCTGCT GTTTTTTAT TTCTATGAG GAGCTTTCTG  
GGAGAGTTT CTGGCTTCA GACTCTGAGA AGCTGCAATT TTATGATGAC TCTATATGAG  
CTGCACCTA CTGGAGAAC CATAGCTGC AGCTTTAGAA AAAGGAAAC CCGGCGCAGAG  
TGTGGGAGAG TGGATAGCA GCAATGCAG GCTTTGAAA ATGAGAGTGG AGAGTTATTC  
CAGGAGAGGT GTAGAGAGAG GATGATCTT GATACATGAA TTGAGATCA TTAGGATCTT  
GTATGGAGCA CTGGCTTCC TCGTTACCTG CTCTTTCTCT CCTCTTGGT GAGAGAGAG  
GCTGCTCAC TCGTCTGGCT TCAATTTTCA GAGCTGCTG CTGCAGTCA AGTTAGCTCA  
TCTCTCTCA CTTTTCTCT TTGCGAAT AGTGAAGTGC ACAGAGTGT GAATGGGACA  
GGAATGCTT TTGATGCAAT CAAAGCTTAA GCTTCTGAG CAGTGTGCT CAGCTGTAAT  
TTCAGTTGCA GCGTGAATGG AAGATTGCA TGGAGCTGA GACTCTCAT GAGAGTGA  
CCTTCATAG AACAGATCC CTGCTTCTCT GCGAAGATC CTTCAGAGGT ACCCCAGCT  
TCCCAAGTT TTTCTCTGT CTGACAGAG AAACAGAG TAACTGATT GCGCTGAC  
CTGGCAGTT GCAATTTCCC TGCAGCTTG ACCCAAGCC AAGACCTGA AAAGGATTC  
AGATTGTGC CAAAACACT GAAAAAACT GCGCTGAGCC TGAACCAAT ACCTGAAAC  
CTGTAAACT CAAAGCTG ACCCCCTGT TTGATATTA CCGAGGTGA ACAGCTCTT  
CTCACTGTCT GTTATGAGA TACCTGTAG CCGACTCAT AATGACATC TCGTAAATA  
TCTTTTGAAC TATACACTT GCGAGTTT TTCTTGGG ACATCAACT TTATCTAGA  
GTATCCAGAG GCACTACTC GGAAGTAAAC ATACTTTAA TGGCTTTCT CTCTCTGTT  
TTAGCTTAG CCTCACTTC CTGAGTTAAC CTCCCAATA CAGATTCAC CTGATCCAA  
GCGCTTAAT TGAAGATAC AGATCACT GTAGCAAGC CTTAGTCA AGCTCTCT  
TGGAGAAC CAACTTAA CAGTCTCT GTCCTCT

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1238

## (2) INFORMATION FOR SEQ ID NO: 87:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1460 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

ATTGCTCTCT GTCCTCTGT GACACTGGG TCAATCTCA TCCCGGAGA GCAATTTCTG  
CTGCTCTCC TGAACCGGG CTGATGGGG GTCCGGAGG CAAATTATC GACATCTCC  
CCACTCTCA TTGCGAAGCT CTTCATGAG GACAGAGCC AGCGATCTT CAGCATCTC  
TACTTTGCA TTCCGATGG CAGTGTCTG GACTTCAATG CAGCTTCCA AGTAAAGAT  
ATGCTGGAG ACTGCACTG GACTTGAAG GTTACCTGG GTCTGAAGT GATGAGCTTT  
CTGCTCTGT TCTGTATAT GCGGAGAGC CAAAGGGAG CCGTGGAGG CCACTCAAT  
TTGCACCCC TGAACCCAC CTCTGATGG GCAATCTGA GCGCTTGGC AAGAAATCT  
AGTTTCTCC TATCTTCTCT GGGCTTCACT GCTGTGCTT TTGTCAAGG CTCCCTGCT  
CTGTGGACTC GCGATCTCT GATGCTTCC CCGATGATCC TTGGGAGAG CCAACCTGC  
CTTCCCGAG AGTCTGATC TTCTCTGAC AGTCTCAT TTGACTCAT GACTGCTG  
ACCGAGTCC TGGTGTGGG CTGATGTGG GAATTCAGC GCGGCTGGG CCACTCCAC  
CCCGGAGTGG ATCCCTGAT CTGTGCACG GCGCTCTGG GCTCTGAC CTTCCTTTC  
CTTCTCTTGT CTTGGAGATC GTGACACTT AATATTTCAT CTTCATTGA  
GAGACCTCC TATTCATGA CTGGGCAATC GTGGCAGCA TTCTGCTGA CCGTGTATC  
CTTACCCAC GCTGCACGC CAGAGCTTC GATATGATC TGTCCACT GCTGGATAT  
GCTGGAGCC CTAACCTCAT TGGCTGATC ATGACAGCC TGCACCGAA CTGACCCCC  
TCTTCTTGT CCAATTTCCG GACTTGGAG TTCTGCTCA TCTTGGCC GTTTGTGGG  
GCACTGGGG GCGCATTTCC TGGCAGCCC CATCTTCAT GAGGGGAGC GCGGCGGGC  
ACAGTGCAC GTGAGAGCC TCTGCAGA AGCAGATTC ACAGACAGC GATTTGTGT  
GCGCAGAG GCGGCTTCA CCGGCTGAC GATGGAGAT GTGCTATCT GAGAGAGCT  
CAGCTCACT ACTGCACAT CTGCAGAGC TGGCTCTGG CCAACCCAC GAGAGAGCTG  
GCGTAAAC CATTGAGTGG GCGCAGTTC CAGAGGAGC CTGGAGCTGG TGCAGCTCC  
CAGACATAC ATGATATCT CAGGAGAGA GATGGAGTTC CAGAGAGGG ATCCCTTCC  
AAAGAGGGA GCGGCAAGG CTGGTCTCA TTGTATAGG GATTAATTT TGTAGCAGA  
AAAAAATA AAAAAAATA

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1320  
1380  
1440  
1460

## (2) INFORMATION FOR SEQ ID NO: 88:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1395 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

5 CAGGTGCAAA GTGGAGATG TGAATCTCTCA GTCTTGGGCT ATTGGGCAC GTGCTCTGG 60  
GACATGGGAC CTGTGAGGCT CAGCAGCTG GAGTCTCTGG CTTTGTGGCT CAGGCTGG 120  
15 AATTGGCCA TTGCCAGGC GGAAGCTGG ACTCAGGCTG CCCCCGGCC GTTCTCTATC 180  
CTCCACCGG ATCTGTGGC GTTGGACTG GGGCTGATG AGTTTCTCA CTTCTGACC 240  
GTATTCTCTC CAGTTTAAG GTAGACATT TCGAGGCCC AGCCAGAAC CTCACCGGA 300  
20 GAAAGTTCAC CCGCGGAGAG CGKCCCGCT GTGTGCTGCC CCGGAGGAC AGCCAGCTTG 360  
TAGGGGGAG TCCACCTGA AAAAAAATT TCCAGTCCC CAAAGGTGA CCGTCTTGG 420  
25 GAGACGGG ATCAGTACC ATGTGGTGC CCACAAAAT TGCACCTTG AGTCTCAGC 480  
TCTGCACCC GGGTCAATT CCAGAGCAA GBACTCCCTC CTGCTTGGA GAGACTTCAC 540  
30 ACCGTATCA CGATCCAC GGTCTGAG GTGGATGGA TTCTGGGTG GATTCATCAC 600  
TCCCGATCA AAGGGCCAA CBAAGCCCA CTAGAACAT GGGTCCCGAG GGTGGGTCA 660  
GGCCCCTTAA ACTGCACT AGTTGGGTG AGCCATTAG ATTAATCTT TTTCTTAAT 720  
35 TTGTAAACA ATCATAGCT TCTGTCACT TATGTATCTT AAGACTCAAT ATTAACCCCT 780  
TGTATTAAT GAGGATCA ATGATTCAT TCCCAGAAA CAGAGTGGG GAGTGTAGT 840  
40 TCCAACTGG TTTTACTAA CCGTGTCTT AGACTCTCC TTCTCTTTAA TCACTCAGCC 900  
TTGTTTCCAC CTGAATTGAC TCTCCCTTAG CTAAAGGCG CAGATGGACT CCATCTTGG 960  
TCTTTTACT GGCAGGCTT TCTTCAGG ACTTAAGT TCCAGCTG CTCCAGGAC 1020  
45 ATCCAGAT GCATTAACT GATAGTAC TGTGGCAG TATATCCGA GTTCCAGGA 1080  
ATTGTCCAA TTGATTAC CAAAGGCC CCGCTCTATC ACCTTGTAT ATCTTAAG 1140  
1200 CCGCTGACC TGGAACTATT AAGTTCCTG TAACCAATTA TCTTTTAAC TTTTGTGCT  
50 ACTTTATTT TGTAAATTG TTTTAAGTAC ACCCCCCCTC TCTTTCTTA ACCAAGTAT 1260  
AAGCAAAAT CTAGCCCTT CTTCAGGCG AGAGAAATTC GAGCTTAGC CCGTCTCTGG 1320  
1380 CCACCACTA AATTAACGA TTCTCTATGT GTAAAAAA AAAAAAAA CTGCGAGGG  
GGGCCCGTA CCCAA 1395

## (2) INFORMATION FOR SEQ ID NO: 89:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1186 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

5 GGCACAGCC GGCAGCCGA GCTAGGTCG AACTGGGG GGCACAGGA TGTNACAG 60  
AAGACAGAA GATGAGACTC TGTTCATTC CTTTCTCTAG CCCATCTTG TGTCTATCTT 120  
15 TCCCCCTCC ATCATACCTC CTCCTTCTG GAGCCTCTGC CGCTTGGCT GTATGTGG 180  
CACTTACCTG GATATTTGAG TGGGAGGATG AAGGCGAGA CTCACCTTAC GCGTGGGAC 240  
20 AGATGGGAG AGCAAAAGG CAGAGATGC CAGGAGGGG GTCCAGACA AACGACAGAG 300  
GTTCGGTCA GGGAAAGGG TGGGAGAAA GAGGGGTGCA GGCCTTGGAG CCGGTGTAGC 360  
CAGCAGCTGC GGCCTCCCG GGCCTTGGC ATCCAACTTC GAGACAGGG TACGAGCTTC 420  
25 CTGTGTGTA TCATAGGANT TGTTCACATA GTGTATGCA TGAATCTCTT AAGGTTAAGA 480  
AGCGGTGTG GTCCACCATG ACATCCACC CGTATATATA AAGATAAATA TATATATATA 540  
30 TGTATCTAA TTATGGCAG AGAATATATA GCACGTAGGG CCCTGCTGCC CTGCTGAGC 600  
AAGCAAACT AAGCCTTTTG GTTGGGTAT TATGTTTCT TTTGTTATTT GTTGTTTTTT 660  
35 GTGGCTTTC TTATGTGTG ATAGCAGAG TCCAGTGGG ATTGCTCTCT ATTACAGAA 720  
AGTGTTTTA ATTCATCAT GTTCTAGTTA ATGTCTACT CAGCAGCTCC TCTTAGCCTA 780  
ATTTAGGAG GTTCCCAAT TTTGTTCTT CAATTTTACT GGTACTTTT TTGTACAAAT 840  
40 CAATCTCTT CTCTCTTCT CTCTCCCA CCTCTCACC TTGCCCTCTC CATCTCCCTC 900  
TCCGCGCTC CCGTCTCTCC TCTGCTCTCC CGTCTCAATT CTGTCACTC CATCTCTCT 960  
CGCTCTCTC TCCGCTCTCC TCCCGCTCC CAGCGCCACT TCCCGGAGT GTGCTTCCG 1020  
1080 CTCTTATCT GTTCTAGTTC CGAAGCAGT TCACTCGAAG TTGTCCAGT CTSGTTCGAG  
CTTTCGANT CTGCTCTCT TTCTGTAGA TTGAGCGTT TCTTTGTAA TTCACTGTTT 1140  
50 CTGACAGAT TTAAAAAA AAAAAAGAA AAAAAAAA AAAAAA 1186

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1821 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

## (2) INFORMATION FOR SEQ ID NO: 90:



(D) TOPOLOGY: Linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

5 AAAAAATGCT TTGAGGGCGT CCGGTATGTA TTGCGGGGGGCG CGAGCGAGAC TGAAGCTGGA 60  
KATCCGTGCT CAGTGGCTC AAGATGGCGT CAGCGAGAC CAGTTACCGA GCTGAAAATC 120  
ACATACCCCT CGAAGATCT CCGAAGACTA GTTGTGAGCA GAGGTGGGGC TGTGAGGCTA 180  
GAGTTGAC ACACATGAC AAGCTGCTGG GGGACATGCG AGAAGCCCTT TTCTGTGCT 240  
CTCCATGCTA GAGCGAGCC CTAGAGAGCT GCGATTACTC CCTGTAGCT AGCAGAGAGC 300  
TGAATGACA CAGTGGCCCA CACAAAAGCA GTGTGATGAC ATCTGTGTTAC GTTGTCCGAT 360  
CAGCGAGAC CAGCCCGCTT TCCGAGTCA AACGAGAGC CGATGCACTT CCAATTGAC 420  
AGTCAAATCA CTACTTCAT TGTACTTTTA GATGAGCAA AGTGGTACT GCTGAGTGT 480  
GTGCGTATCC GTAGAGGCG CAGCGAGAGG ATGCCCAAG CCGAAGCTTC TCGAGGGCTG 540  
CAGCGAGAGC AACCCGACA GCTGATGTC AGCAGAGAGC CTCCGAGCCA ATGTCTGTGT 600  
CTAATTTGAA GGGGAGTTTA GCGAGAGAAA ATCAACTTAT GTATATTTAC AAAATCTCT 660  
TGACTGACT TAAAGTTCT ATGAGACT ATCTGTTTAA AACCTATGG AAAATCTGTA 720  
TACTTAAACA AATGAGCTC TAAAGAAAT TTGGAGGCCC ATGCTGAGAA CAAATCTGT 780  
GAGTGAAGAA TGTTCGAGA AGCTACTTAC CTACATGTA ATGTGCAAT TTCTTCTCT 840  
TTGTAGAAA AATCCCTCTT ACTTTTGA ACGATATGG CAGCTCTAG TGAAGCAAT 900  
ACATGTTGAT ATGAGAAAA TTAAGATTA GTATAAAT GTTAAATAT CCAATATGG 960  
ATATGATGG CGAGAGATT TAACTTACA AGTATTTCT AATGTAAAG TATTCAGCTC 1020  
TTCCAGGTGG AATCCCGCT ATCCGAGCT CAGCTTCAC ATCATTTCA AAAAGCAAT 1080  
TCTGTGTC CAGTATCT CCAATAGCT AACTTTTGG TCCGTCAAT AGCAATTT 1140  
AAATCTGCA AAGATTAAT ATCATGAT TATTTAGCT TTACAAAAA AAAATGTAAT 1200  
GCGCTTTTAT TTGATGGCC TATTAAGAG TACCTTACT GGGCAGATAT AGGAAAAACA 1260  
AATTAGACA AAAAGAAATC TGTAAATC CAGCGAGAA AAAATGCTGG CAGAGTACT 1320  
CAATTCCTGG TCCCTCCAT CAGTCAAAAT CAGAGAGCTG CAGTGAATCG CTGTCTTGG 1380  
AATGTGAGC AGTGTGCT GTAGCTTTT AAAATGCGC TATTAAGCTT TTACAGCAAT 1440  
ACATTTAAA GATTAAGTTA CTTAAAGAGC AAAATTAAT TATAGTCTT CAGCAATTTA 1500  
TGAATCTTGG GTTGTCAAT TTGTTCAAT AACTTTTCT ATCAAGTCT GTTGTCTTGG 1560  
TAAAGATTC TCAATTTGG GTTGAAGG TAAAGCAAA CTGTATTTTC AGTGAAGAAA 1620  
GTATTTGTA TGAATTTTA AACCAATTA ATGTATAAA GGTATAAAAA AAAAAAAA 1680

AAAAAAAAA AAAAAAAAAA AAAAAAAAAA CTTGGGGCGC GAGCTTTTCT CCGTTGGGCT 1740  
GAGGGTTTAT TTGAGCTTGG GCGACTGGGC CTTTGTGTTT TACAGCTGGG TGAAGGGGGG 1800  
5 AACCGGGGGG GGGTTTCCCC C 1821

(2) INFORMATION FOR SEQ ID NO: 91:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 862 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: Linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

20 TCCCTTTT CCGACCAAT CCGGGGATGG TGAAGTGGG AAGTGTGAC TCAATTTAG 60  
GAGTGAAGA GAGTGAAGA ATTGTGAG TGGAGATTT GATTTGAT TTGAGCTGT 120  
AAATGACTGG AACTTGCAAT CTGTGTGTA GGTGAGGTT TGTGTGGGG TTCTGTGAG 180  
AGCTTACTGA CCCCCGATG TTTTCTTCT CTGTGTGTA GAGAGCCCT GTTGTGTCT 240  
TACAGCGCT TGAAGTCTC CGAGAGACA AACAGGAGA GAGGAGCTG TAGGAGAGT 300  
TCTTCTGCT TTTCTGTGCT TTCTTTTAA CAGAGCTCC GAAAGCGAC TCAATGCGAT 360  
GCGAGAGCT TTCTGAGAA CAGTCAAAA GATCTGTT AGTCTTTCT TTGTCTTCC 420  
AGTGAAGCT TCTTATCCA CCGTTCTGG TGTCTATGG AATCATGAG AAGAGCTGG 480  
GAGTTTTC TGTCTCTTC TGGCCTTCA TGAAGCAAT GAGCTGGGCG TCGCGGGCTC 540  
CTGAGCTGCA CAAATTAAT CTGTCTCCG TCCGAGCCCT TCTTTTGTCT CTGAAAGCG 600  
TTTAAATG TGAAGTCCC AAGAGAGAG CCGCTGAGTG TATGAATTT GAGCGCGCTT 660  
TTGTAGGCT CGAGCCCCA AACTTAAGT AACTTAACA ATTTTAAAA GATTCATG 720  
CTGTGATCT CTGAGAGAT AGTATGAT GTAGAGTGG CAGCGAGTG TGTGAAATG 780  
45 TGTGAGCTC GATTTATAG ATTCAATAT TAAATGTCT GCTGATATA AAAAAAAA 840  
AAAAACTGA GGGGGGCGCG GT 862

(2) INFORMATION FOR SEQ ID NO: 92:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 696 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: Linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

CTGAGGGCAG TGAAGTGGAC TCTGAGGCT ACCCTACCG CCAGTGTCTCC GCGAGGGGG 60  
5 TGGAGGGCAG AGGGCCCGCG AGGCCCGAGT TCCAAACATG GCTCAGAGCA GAGAGCGCG 120  
AAACCGCTTC CCGAGGCCCA GCGAGCTTGA CAACCCCTTT CAGGAGCCAG CTGTGATCCA 180  
GACCGACCC AGCCGGCAGT ATGCCAGCT TGAAGCTTAC AACCTTTTG AGAGCCGCGA 240  
GCCACACCA CCGTATGAGC CTCAGCGCC CTCGCCATTT CTTCCACCT CAGCTCCCTC 300  
CTTCCAGCCC TCGAAGAAC TGAAGCCGAC AGAAGCTTAG AACTATGCT CATAGAGCAC 360  
15 TCAAGCTCTA CTTCCAGCAG CCAAGCTGA GCTCTGAG GAAAGAGAG AGCTCAACCG 420  
GAAGCAGAG GAGTTGGACC GAAGAGCGA GAGCTGAGC ATGCTGCCCT GGGGGCACA 480  
GCTACTCGAC AGAACATTTG GCGCCCTCTA CTTCTTTT GTCCAGTTCA GCGCTGCTTT 540  
20 TTCCAGACA TCTCCATGGA GATCCCGCA GATTTTGA AGACTGTATC CAGCATTAC 600  
TACCTCTGGA TGTGAGGAC GGTGATCTT CTCTGAAVT TGTGAGCTG CTGCGCCAT 660  
25 TCTGTGGA AACCAACAT GCGAGGCTT TGGGTT 696

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1886 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 91:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

40 CAGGCCACTG AGCTTCTTT CCGAGGAGT CAGGAGTCC TACAGAGAA GCGCTTCTT 60  
GCATTCAGA GGGCCACAG CTTCTACCC AGAGTACCC AGCAGCTTT CTAGCTGCTT 120  
CTCAGCTTG CCATCTCAG AGAGTCCA GAGCTCTCG GGTATGTCC CCAAGCTTT 180  
45 CAGCTTCAG GTGAGATGC CAATCCCTG CAGCTCTTG CCGCTCTCT GTACGACAG 240  
AAGCATTACC ATGAGGCTCT GACATCATC GACATGCCC TGAATGATA CCGAGAAAT 300  
50 TTCTACTAC TGTTTTCCA AGTGAAGTG CAGTCACTCT GCGAGGCC GAGAGAGCA 360  
CTCTGACTT GTAGACAT GCTCCAGTA TGGAACTCT GCTACACT CACCAACCC 420  
AGTATTTG GAGCTGGAG CAGCTCTTA GATAGAGCA TTCTGACAG AGCAGCTT 480  
55 AATACATTA CTTTCCAGA CTTACAGAT CCGAGAGAG GCTCCGCTCA TCCAGATG 540  
GTAGCAGGCT CAGAGTGA GCGGCACTG TCGAGATG CTTTCTCTT CAGAGAGCTG 600  
60 CCGCTTAGCA GGGCCGCTG CAGCCCTGGA TGAAGCTGC ACAGATCTG CTCATGAG 660

CTGAGCTTA TATCGGATC GCGAGGCTG CAGAGGCC AGCTGTACC CAGAGGCTG 720  
CCAGCTCTT CCCATGTCC CACATGTCC TCTACATGCG GCGCAGATT CGTGAGCTC 780  
5 GCGAAGCAT GAGAGAGCG GCGCGGTGT ATGAAGAGG CTTAGCCAT CAGCCCCAC 840  
CAGGTGAGA GATGCAAGC ACTTGCCCT GATCCTTAC CAGTTAGCC GTTACAGTT 900  
GGCGGAGAG ATCTCCGGG AGCGGTGCA GGTCAACTG ACAGCCGAG AGTCTGGA 960  
CGCGTGGG GAGTCTCTC AAGCTCAGG CAGCATGCG GCGCTACCG AGTCTTCTT 1020  
GACAGCTTG GAGCTGAGG CAGCAGGCC GCGGTGCG TTCACTACA TCCGCCGCTT 1080  
15 GCTCTGAGA GCGGCTGCC AGCTCACT GCGGCTCAG CTNCGAGGC CTTGCCGGC 1140  
ACAGGCTT GTCCATGCG CCGAGGGGA TGAATCTGC GCACTGAGC CAGGAGCAG 1200  
TGTTCAGTG GGCAGAGTA ACCAACAAA CCAACCCGA ATCATGCTC TCGCATGTTG 1260  
CGTTCTCTT GTTTTTTTC CAGGCCAT GTTATGTTCT GAACTATTG ACATGTTCA 1320  
AATGATCA TGTGCTAT TTTTGTAGT GACATCTG TTTTCTAGTA AATGATTTG 1380  
25 GATTTATCA GCAATGTAG AAGATATAT TCAAGTTTA AATTCAGTG CAGCAGAT 1440  
TATTTTATC AGAGCTGTA AGAAGACAC TGTCTTTTC TCCCGACAC CCGCTCTGC 1500  
CCACTTTGC CAGAAACA AATGTAACT TCGTGTCTC CACCTCAGCA CTAGTCCATG 1560  
CAGGACACC AGCTGACAT TTTTGTGTT TACTGTCAAT AATGTACCA TGTGATCAAT 1620  
TACTCTCTC ACTTAGACA AGCTGTAGT CCGAGATAT TATTTTITA CCAATATATG 1680  
35 CCGTTACAA GAGAGGAAA TATGATTTT TTAAGTTTA CTTTTTTAT TGAATTCAGA 1740  
GTTTATTTAT CAGGGAAT ATGTACAAAG AAGCTTCAA TGAATATTT ACCGACATTC 1800  
CTTATACATG ACAGACTCT CCGTACATG GAGATGATG TTATATATTA AATGATTTT 1860  
AATCGAAA AAAAAAAAAA AAAAA 1886

(2) INFORMATION FOR SEQ ID NO: 94:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1774 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

CTCAGCTACC GTATACAGTA GAGATTAAC CCAATTCACA TCCACTAC TCGAGCTTCC 60  
CTCCTCTGCC CCCCATTTGA AGATGTCTT TTTCTATAC TATATCTAT TCCATCTCAT 120

GAATATCTCG TAAATTAATT AATCCCTCAT GAAATGATAA TTAAGTTCAT TAAATATAGA 180  
AGTGTAAATTA AAGATTCCTGT AAGATTAATT TCGTACTTCAT GTGGATTAATT CTGAGAGAG 240  
AATAAGTACA AATTATATCG ATCAGATTCG ACAATTCGAG TTGTGAAAC TACTACCAAA 300  
AAGATTCAG CAATTAACA CTCGATCATT AGTAAAGAGG CCGTGTTCGC TAAATCTGC 360  
CAACCTGAA TCGTTAAAAA TTGTGGGCAA TCGGTAGGC AAAATATCTT TCTTTCTTT 420  
GAATTAAT GAGAGGAG ACCTTCAT GTTTCGTGC CATTGCAT TCGTATATG 480  
AATTCCTTT GCGCATTTT CTTTTAAA TTATGAAGT GTAAATACA CCGTGCAT 540  
GTAAAAAAA GAGATCTT TGAATAGAG GAGCTTTTC TCGAATGTA CCAATCAGAT 600  
TCCACTTAC AAGTTTAC ATACATCTC TAACTACT TCGGTAGAA TATACATACA 660  
CATAAAAA CTTTTCAT AAATAGATC TGAATCTCG TAACTTTTA AATTTTCGT 720  
CTCAAAAAA GATACAGGT CTTTAAATT CTTTAAATGT TGAATATAT TAAATATAT 780  
GAAATGCA TTAATTAAT CTTAAATTT TTTCCTCG CTATACAT GCCAAAGTAA 840  
AATCTTAAT GAAATCTT TGTGATGTC TGTAAATAT TCTTAAGCT GAGTGCAGT 900  
AAGGTGAT TTGTGATCA AAGATTCAT CTCGTGAC CTGAGTCT CCGAAAGGCT 960  
TTCATAGTC TATTTTAC AAGTATAG AAATATGA TTCCCGATA TAACTTAC 1020  
TAACTTCAAT AATTTTATC TGTGGGCA AAAAGACT TTTCATTT TCAATTTCC 1080  
TGAATTAATA AAAAATGT GAGTTGGG TTAATTTCA GTTAAATGAC CATTATAT 1140  
TAACTGTA TGTATTAAT CCGTACCTG CTTTCAATG GATTATGT GAAATATG 1200  
TTTATATG TTCAGATCT CCGTCCCAT CTCTGATTA GAGACTTT ATTATTTGT 1260  
GGGAACTTA TTCCCTGCG CTAGGTAG CATGTAGCA GCGCTGGCT CCGAGTCC 1320  
AAGCTTCCT AAGCAAGTC AAAAAGAT GGTATATTA CTTAAGCAG GTAAAGAAA 1380  
GCCCTTACA GAACTTCGC TGAATACT GGAAGAGAG CTTATAGAG ATCCAGAAA 1440  
CGAAGACA TGTAGCTG AATTGTGCG ATGTGAGAG AGTCTGTG AGAAGAACT 1500  
CGAATGCGA CAGAAATGGA AAGAGACTA AGTCTGATG TCAATTTCT GAGAGCTTA 1560  
GATTCAGTC TGCCTAAGC CTGCGCATC CCGACTTAA AAGTTTGTG AGCAATAAA 1620  
GTCCCTTCT TGTTAATAT AATTGAATG AGTTTCTGT CTGATTAATA TAACTATTT 1680  
GTATTTCTT AATGATTTGT AAAAACTT TGAATTTTA AATTCAGAC TTATGCACT 1740  
AATTAAGTAA AAAAATTTAG CATGCGCTC CAGG 1774

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2503 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

GGAGAGAGC AAGGAGAGC GAGACACT GAGACTGCA TGTGCTGCC ATTTCCTTC 60  
CACTCTCT TTCTGATGC TGAATTAAGA AAGGAGAGCA GAGAGAAAT TGAAGAGAC 120  
AAGCTTAAT TCTGCTCT CCGTTTATG AGTATCTCG TGTCTCTGC AGCTCTTTC 180  
TGTCCCGCG GAGAGGAGC TAAAGATGAG GAGGCGAAG CCGTGTTCGC GAGACAGACT 240  
GCTTTTATG AAGTCAATAG TCTCTGAT TGTGTATC TCCGTAAT TAAAGCTCT 300  
TCTGTAGAG GCTGAGAC TCACTGACT GCGCAAGTC AATATCGGCT TGAATACTT 360  
CTGCTGTG AATGAGAGCA CAGACACT AAGATTCAC CAGTTCCCTG AGCTGAGAC 420  
CTGAGGAGTC GCTGAGATG GCTGAGCT GAGCAGGCTT GAGTGTAGC GGTCCCTGCT 480  
CTGAGCTTC TTTCGCGCCC AAGCTCTCT CCGTACCGAG TCGAAGAGTC AAGAGAGAG 540  
GTGAGGCTG GAGATGAGCT TCTGCTGCT ATCTCTGTG CTGCTGAGAG GCGGCTGAG 600  
CGCTTCCTC TCGTATGTGT GGAATGAGTC AAGCTCTCC TCCGAGGCTC TGGATTTCTA 660  
GCTGTGAGCA GCGGCGAGGC CTTAATATC CTCTTCTTA TAAAGATAGC TGTGTCTCT 720  
CTGAGGCTG AAGAGGCTGA GAGCAGCTT GAGAGCTGT AAAGCTTAC GTGATTTGAA 780  
GAGTCAAT TCAAGCATG TGAAGATGTC CAGATCTCT GAGCATTC ATTATTAAGC 840  
TAACTCAAT CTGCACTTC AAGATGAAA CCGATTAAG AGAAGATGAG GCGCTGTGT 900  
GAAAGAGTC GAGGAGAGC AAGGAGAGC AAGGAGCTG TCACTTTTA GTACAAATT 960  
GTCTGCTCT CAGACTTTC AAGCTGCA AAAGCTCTT AAACATGCA GCTCATGTC 1020  
AAGCAATTC CTGCTTAA TAAATATCT GAGCAATCT TCGTCACT TCAAGAGGCT 1080  
GGGAGAGAG TGAATCTGT CATGGAGCA GACTTCAGC CTGATTTGC AAAATGAAA 1140  
AATGAAGTA GGAAGAACT TACGGCACA AAGAGAGCA TTAAGAGAG GAGACTTCA 1200  
GTGTCTGAG GAGATGTGT AAGAGCTTC CACTGACCC ACGATATGA GTGAGCGGCT 1260  
AATAGCATC ACTGAACTC CAAGCCAGA GCTCAAGAG TGAATCTGA GTACATCAC 1320  
AAAGAGAG CAGATGAGT ACAAAGACA GCTTGAGTC CTTCAAAGCT TCCGTGAGC 1380  
TGAAGCAGA GAGGAGAG GCTGCTCTG ACAAATCACT CCGATATGA GAGCTGAG 1440  
GAGTCAAT CCGCTGTG AAGAGAGTC GAGGCTGAG CAGAAAGAA GCTGTGAGC 1500  
CTGCATTC GAGCCAGAA CTTGAAAGA CATTAAGCA ACTTAAGAT TGGGATTTG 1560

5 TGTGTAATTTT CTAGCACTTG TGTATTGGAA AACTGTATG CAGTGAATTT ATTCAATAT 1620  
TCTGTTCGAA AGCCACACTG AAMACAGG CAGAGACATG TACTCTGGTG TGATCTCTTG 1680  
TCTCTAGTGT CTCTCTGGG CTCTGTGCC TCTTCTTTTA TAGCTAGCTG CCGGGGAC 1740  
AAGGTACAGG TGAAGGAG GTAGAGCTT GCGGGAGGAG GCTGTCTG GTTACAGCTC 1800  
10 TATACACTGT GGCCTCAGC TCCACAGAG GGCAGAGAAC TGTGGGAGG TCGTTTCTT 1860  
TCTAGGCTGG CTGGAGAGT GGGAGCTCAT TGTATGACTC ATGATGGAAA CTATTTTGA 1920  
AACAGGCTTC CTCTCTCAGG AGAGATCATG GGGACTTAAC TGTAGCAATT CAGTGCACC 1980  
15 TGGAGTGAAT CTTTTTCTTT GCAAGTACT GTCTCTTGG TTCCAGTAG TTGGACACC 2040  
ACATGACATY ATTTTCCCTG GAACTGGTC ACTGACTTAC AGAGACATTT GGGACTCCAG 2100  
20 AGCCTCAGG GCGAGGAG GGCACAGTAC ATACAGAGG AGTCAAATG GATCTCATTT 2160  
TGAGTCTCTG CTTCCTCACA CTCAGACGG CAGCCCAAG GCGCGAGTG TCCAGGGCTT 2220  
25 CTGGGCTGAG GTGATCTGC CAGGCCAAG AAGGCACAAA GTTAGAGACA CAGAGGCC 2280  
CATTCACACA GCGGKGCG CCGAGCAGC CAGTGGAGC TCAGCTCTCC TCCAGCTGCT 2340  
CTCGGAGAC AGTCACTGC ACAGTTTATG CCTAGCTCA AAGATCTC CCGAGCTAT 2400  
30 TTACAGCAT CTTTCTCTC CTCTCTCA GGGCTCTCC TAGAGGAGA GCTGGAGCC 2460  
CCGCGCTCT GGGAGGGCT GAGGCTGGA GTCAGTGGCT GTC 2503

(2) INFORMATION FOR SEQ ID NO: 96:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2801 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

5 CTGGAAGCC GAGGTAGCC GAGCGGGG GGGCTCTCG AGCGCGGGT GCTCGGGTG 60  
CGCTCGCTC GCGCAGAGC ACCGAGAGC GAGCGGGG CCGCGCGCC TCCCTCTCA 120  
TGAGGCGCGA GTGAGCGCG GCGGCTATG CCGACCCCG GCGCTTTCC CCGCGCTCT 180  
ATCGGAGCG CAGCAGAGC GCGCCCTGCA GAGGAGGGG GAGGAGAGG AGCTGTGCG 240  
55 AGGTTTCA TCGGCCCCA GTGCTGGTC GGGCGCGAR CCGCGGCTG GCGCGCGAG 300  
GGGCGGGCC TAGGGCGGC GGTTCGGA AMGARGGR GCTCTCTAG CCGCGCGCC 360  
60 ACCAGCGCC GCAACDAA GCGCGGGGCT TGTCTGACC AMCGCGGCTG GCGCARCCA 420

GGAGAGCCC GCGCGAGGG GCGCAGTGG AGGTCCCGC CAGCCCCCG GCGCTTCCC 480  
GCGCGGGCC GCTCCGAGG CAAGCGGGC CCGTGAGAG TCCACTAGT TCACAGATA 540  
5 AAATCCACA CGAGACTCG GAGTCAGAA TGGCTAGCC CCAAGTGGT GTAGCTCTG 600  
TATTAATGC TAGCTGTCT GTGAATGCC CTGAATTTTA CCGTCAAGT TATTTCTCA 660  
GTACACAGA ATCTATAGG GATGTTGAG AGGATTAATC TACTCTATCA GATATGTC 720  
780 AGGATTTTT GAATCATCT ACAGAGAGC CTGGAGTTT TGAAGCTGAA ATTGAACAGT  
TTGAGAGAC CCGTAATGTT TGTGTTACA CAGATATGC TTTCAGAAA CTTGTGAGC 840  
15 TCATCTATCA ACAGGCCCA TGTATCCCA ATTTCTCTTA TATGGAGCT GCGCTGTGA 900  
ATTACTGTC CCACTATCTG ACAATTAGCC CACAGAGTG CAACTTCCG CAATTGCTAC 960  
1020 TTCAAGATG TGGGACTGAA TATGAAGTTA AGATCAAGC TCGAAGAGG GATGAAGTTA  
1080 CTGGAAGG ATTTCAATCA TTGTACTCT TTCTGGGAGA ACTTTATCTT AACCTGAGA  
1140 TCAAGGAGC AATGAGAG GTTACAAGC CAGATATCT TCAAGTTGCT CTTGAGAGAT  
1200 TCGTGAATC CCGTTTCTT AATCTATG ATGACAAATT AATTGTGCA GTAAATGTT 1260  
TAAAGTTGAC AGATCAGTT TTGGAGATG CTTCGAGGA AAGAGGAGG ATGATATG 1320  
AAGAAATAT TCAGAGATT GAAAGCTG TCTAGATGC AACTGCACT AGAGATGTA 1380  
AACAGATCT CTTGAAGCT GTAGAACTC GTCAAGTAA CTGGGGAGA GTCCATGCA 1440  
CTTCACATA TAGAGAGCA ACACAGAAA ATGATCTTAA CTACTTTATG AATGACCCA 1500  
35 CATTATATC ATCTAGTGT GTCTTTTCA CTCAGCTGA TCCAGATTAC CAGAGAAAT 1560  
ACCAAGATT ACTGAAAG GAGGACTTTT TTCAGATTA TGAAGAAAT GGAACAGAT 1620  
TATCCGGGGC TGGTGATCCA TACTTGGATG ATATGTATGA TCGATGGAC CCGAGATAG 1680  
AAGAGCTTA TGAAGTTT TTTTGGAT CAGAGCTTA GCGAAGCAG TAAAGTTAA 1740  
TTTCAGATA TCAGTTTAT AAGCAGTTT AGGTATGTTG ATTTAGGAGA ACACAAGGA 1800  
45 GCAAGAAAT GTGTACATC TATACCAAT TAGGATGTT GAGTATGTT ACTAATGAT 1860  
GCACTTTAA TTTTGTTAA CACTATCTG CAATATAAC TTTATTCCT ATACTTTAA 1920  
ATGTATAT ATATATAATA GTTTATATG TACAGTTAAT TCTACTGTT TGGCTGCAAT 1980  
50 AAATCGATT TTGAATAAA TGAATGTTG AATTTTTC TAGTTGTTA GATGCTTATC 2040  
CTTTAATTC TACTTTCTT GAGGGGAAA AGCTTGTGC TCGAATACA TATTAATGCA 2100  
55 AAATGTAGC ATCTTTTTT AGTATGAGT ATTATAGCTT YCATTTTACT TKGACATTA 2160  
GTGTCCCAAT GAATGAAT TCAATATGA ATCAATATCT TGAATATCTT TAGCACTAA 2220  
GTCTTGGGA TATATCACA ACTGATTTAC ATATGAGAT GCTATTTGUA TACCAAGGC

TTTTTAAGG TGAAGGAGG GAAAGACCA ACTTGATGA ACTCCAGCT AAGACCA 2280  
GACTCAGC GAGATTAT TCAATTGA GGAATGTC TTTTATTI TTAATTTC 2340  
5 AACTGCTA CTTCATAC TTAAGCTA GAACTATT CTCTTATT TAAATGCTT 2400  
TGCACCTTT ATTTCAGC AAGGATGA TCACTTTT TACTAGAA TTTTACGA 2460  
TATTTCAT TCAAGTTTC ACTGCTTA TATTGAGC ACTGAGGG TCACTTTCT 2520  
10 AATCTTTA CTATATGTC GTACACCTC AACCCTATA AATATATT TTAATATG 2580  
TCAATCTTT TCCAGCTA CTAAAGCTG TGTACAAAG GATTCCTGT AATATGCA 2640  
GTAAATGAT CTGTATTA CCACTGTT TACATTGAT ACATCTGT CTGTATAC 2700  
15 AATTACCTT CTCACCTTC TCTTGTTG TTAAGCTGA ATTAAATTA GACTTGAA 2760  
20 ATTAAGCTTA AAAAAAAAA AAAAACTGA G 2801

## (2) INFORMATION FOR SEQ ID NO: 97:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1631 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

ATGAGCCA AGACATAC TGAATCTTG GCTTCTAGTA TAAATAGG TGTGCTCT 60  
35 AATTCTTC CATACATAT CATGCTAC AGTAGATCT CATGATGTA ACTGCTCTC 120  
GACTGCTTC TACTCAGAT TGTCTGCA GCAATACGA AACAGAGCA CAGAGAGCA 180  
40 TGCATAGAG GCTGATGAG AACTGAGCT GTAGACCCCG GATTAATCT GAACTTCAT 240  
TCTTATTAT TGGAGACCA GAGAGAAAT GAAAGAGTC CAATCAGCA AGTTACAT 300  
AATCAGATG CTGAAATTA CAGCTTAT CCTTGTGTG GAGAGGCTT TCAATCAGCC 360  
45 CACACAGCA TACTCAGCA GAGAGGCTT GTTGATTTTC AACTTACCG CCACTTTA 420  
AATTTCAC TCAAGATAT TCTGTGAT GTCTCATGT GTATACAT ACTGATGCC 480  
50 AACTCATAT GCTTACTTT ACCAGATTT GCTGAGCTT GATTATGTC GTTTGAGCG 540  
GGAAGTCGA AATTCATGA GCTTACAGA GCTGCTTGTG GTCTCATGT TTGCTGCTA 600  
ACCATAGAG CTGTAGAGT GATGTGACA TGAATGCTC AAGACACAG AGTAGATTC 660  
55 CAGAGGTGA AAGATGCTC TCTCATATC ATGAAGACT TAAATGTCG GATGCTGTC 720  
GCTGAGTGG TCTCTCTCT TCTGAGCTC CTGTTAGAG TGTCTATGT GCTGCTCTG 780  
60 AAGATCTCT TGAATCAGC TCTCTTTT TATTCATGCG AAGACTGAGC ACTGAGATC 840

CTGATCCA AATCATGCG AACTTACGA TTGATGCTC CTGATGCTG GTTGAATCT 900  
GTATTCAGC AAGTTAGAG AATGAGTC CGAGATATG AACTTACGA TATGTTGCT 960  
5 AACTGAGC CTCCGTGAT CTCTGCTG TTCTTTCC TGTGTAGC TTATGTATA 1020  
GCTTGTGTG TTGCTCTT ACTAGGTT ACTGAGAA TCAAACTT ACTCATGCG 1080  
10 GGAATTATC CATTTTACT GATGCTGTC GTATGATG GATTTTTC CTTCAGATC 1140  
CGCAGTTTA AAGCTTTA TGAATATT AAAAAACA AGTACTGTT GGTCAACGA 1200  
CTGTGACT AAGACAGAA ATCTGCAA CAGCTCAT CTGACAGC TCAAGATCA 1260  
15 TCCAGATAT AATGATGTC TGTACAGC TTAGCTTCC GATTATAG TCTTTTTC 1320  
TCACTCTC TCTTGAGA TTTTCCAG TGAATCTGA GCTGTTT TAACTTAAT 1380  
20 GTATTGACT TGTGTCTCA GATTCAGAG AAGAGGTC TAAATGTC CTGTCTCC 1440  
TGAATCTCT GACATACG CTGTCTAGA TTTGTATAG TGTAAATCA AGTCTCTGA 1500  
25 TACCTTAAA CCTGTATTA AACAGATGT GCAATACGA TCTTAAACA AATGATAT 1560  
TAAATTATTA AATCATGTC TCACTTAAA AAAAAAAAA AAAAACTG AAGGAGCTCC 1620  
30 GATACAAA T 1631

## (2) INFORMATION FOR SEQ ID NO: 98:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 504 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

CGAGTGGG CGAGATGAG GGAAGGAC GAGCGCCCG GATGAGGTT GATATGCTT 60  
45 CGAGAACT ACTGAGCG CAGCTGAG AAGATTAG GAAATGTC CTGTGAGTC 120  
TCAAGAGCG ATGATAGAG TCAAGCGAA AATTAACAT CCCCCCTT GCTTCAGAT 180  
GAAAGGAC GTGAGATG TGGAGTGA TATTACAC TCACTGTAA CACAGATAT 240  
50 CATGCTCAT GTATCTGAT TGGACTGAT ACCAGAACT ACACATTA GATGTGTG 300  
AAGGTGTTT GCACTGTGA CAGCATGAG CTGTCTGCC GAGGAGGCC TTATTCAGG 360  
55 GAACTCTG TTCAATCCA GCGATCTTA CCAAGAAA CTTGTGATG AAAAAAGA 420  
AGTTGTGA TTTTATTA GTTTTATG TGAATAGAG TATTAACAT ATTCTGAT 480  
60 TCTTCAAAA AAAAAAAAA AAAA 504

## (2) INFORMATION FOR SEQ ID NO: 99:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1416 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

GGCAGAGGG AGGAGCCCT CTCGTTGGG TGACTCTTGT GTGCCCTTTA GACAGGCTGG 60  
CTGCGCGGTT CCACAGGTTA CAGTAGAGAC TTGAGTCTTT CTTTTCTGT TTGAGTTGG 120  
TGAGTGAGTG ATAGGTTAC ATGGGCTTC AGGATGACC CTTGGAGCTG TCCGAGTTC 180  
CTTAATCTC AGCTGGATC CTGAGCTGG GAGCGCCCTG TGAAGGCCAG CTCGGANAA 240  
ACCTGGAGT TGAATCGGA CTTGTGGAG AACTCTGCTC GAGGGCAGGG TCCCTGGAA 300  
CACTGGTACT TCTGGGCTG GAGGGAGAG GGGCTCGGC TTTCTCTGAA ATGAACTG 360  
CTCTCAGCA GTTCAAGTAC TTCTCTCAA AACATTTCT AATTGATTG TAGGTTTCA 420  
TAAGCATTT TCTTTAAG CATGGAAG GAAGATGCT CAGGAGTC ATGTTTGT 480  
TCACTGGGAT GGGCGCGT TCTACTGCT GGGGCTTCC CTTCTATG GACCTTTGT 540  
GAGGGGCA CCAAGCAGAC TCTTCCACC TTCTCCACT GAGCAGCA GGGGCTTGA 600  
ACCTTAATTT GCTATATCAG AGGATTTTT TTTGCTCTAG TATCTTTAC ACTGTGCAA 660  
CGCTCTAAT TTTTAAAG TTCTCTGCT TGTATTACA GGAATAGA GAGAAATAGT 720  
TTCTGAAGC AGTTTATTT GAAGATCCC AAGGGAGGT TCGTAGAGA AAAATAGTAA 780  
GCTGGTTAG AACATGACA GGGAAACAG CAGAGAGCA TTGGAGAGA ATTTGCCAA 840  
GATCTACCT GAGATACCC CTGTCCAGTG TCTTCACAC GTGATTAAC AGCGTCCA 900  
AGTGTTTTC TCTTTGAAA AAAAAATTC CACAGCTTT TAAAGTGCA TTTAGAAATC 960  
CATGTGACT TAGAATGAAA CTGCGGCCC TGGCAACTGT CAGCTGTGCT AGAAGTTGG 1020  
ATGCTCTGG ATGCAATG ATACTCACT CATTTGTT TCTTGATG CATTTTGTT 1080  
CTTTAGCAG ATCTGTCTT GTGGTGCTG TCTAAGAGT CGGACACTT GTTTTTGTG 1140  
TTAGATTGAG CTGGGAGCT GCATCAGCT TCTTTATNG CAATTAGGC AGGACCATC 1200  
TGTGCTTCT GTTGTGTGC TAATGAGTG AGGGAGGGA GGAATGTAC CCCAAGATA 1260  
GGCCCTCCA TTGGCTTGG CCAAGCCCA CACTTCACAT GTTTACATG GTTCTGTGA 1320  
ATTTAAAGT TTAATGTAT AAAGCGAGC TGTTCCTGTG AAAGTATATA TTTTGTAAAT 1380  
AAATATATTG CTACTTGAA AAAAAAAAA AAAAA 1416

## (2) INFORMATION FOR SEQ ID NO: 100:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2847 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

GGCTAGGACA ATTTTGCTGC TTACTCTATC TCTGCANAG CTGGAGANT TGGCATCCA 60  
TTAATTACAA CCACCAATCA TATCCACAA AAGTACCCTA AAGAAGGAC CAGTGGCCAC 120  
TCTCGAAAA ATTTAAGTAT CAGAAGATTA AAAGATTTT AGGATTTGGA AGCTGTATTT 180  
GTCTTTCCCC AATATCAATT GTTGATCTC CAATAGTAG CTTATATTA GCANTGACA 240  
GATCATGGT TCTCCATATC TCAATCATAG TTACTACTTT GGAATCAGTA TTTGGCANA 300  
TTCAAGCATT TATCAGTGG ATATNAATGG AAATATNAA ATATTTCCA ACCGTCTCA 360  
GTACTCTATC ATATCTCTGT GATCTCTCAA GGAAGGACT TTTGCTTTTA CTTAGAAAGC 420  
GTTTCAGATT TCGTTATAG ACTCTGCTG TCTTCAGTAC CTGATNAAC TTAAACGAG 480  
GAAGCATTA ACAGATGCA GCAAGTTTTG CCAAGCTTC TAAGTCTCG CCGCAGCAT 540  
TTATCATGTT AAGACTAGG ATGCTTCTG CAGTGGCACT ACTTTCCTT AGAGCTGGAG 600  
CATGCTGCTT GGGCTTAAGC CCAAGCATGA TGAAGCTTCC CTCCTGCCAG GTCAATAAA 660  
GTTAGAGAGC TCAAGATTGG GTCTTCCCTG GGTCAAGTG GCAAGGTTTG CTGAAACCCC 720  
TAAGAGAG TGACCAAGGG AGGAGGTAA TGAATGTTTC CAGATCACT CAGATCTCA 780  
TAGCAATTC TGGCTACTT TCAATGTTG AATTTCTGGA TCGTAGAGG GACTTTGATT 840  
TGATATAT AAATCCAGGA CAGTCCCAAG AAGTCTTGG AGTCTGGCT CTGACAGCCC 900  
AAGAGGGA ATACTTGTA TTAGGAGCA ACTATAGCC AGCGCTGAG CTGTCTTTA 960  
GATATAAA CAGATGGGA GTGGAGAGT CATTGCTTC AAGTTATACA GCTAGGAAT 1020  
ACTCAAGCA AATCTTGAC GCACTCCCC CTAAATCTGT GGAAGGAC TTGTACCAAC 1080  
ACACCTGCT CCACATAAA ACAGAGGAT AAAGACTTT CAGGTTTCC CACTGTGTC 1140  
TGACCATCC AATTTATGA TCTTCTTCA ATCAATTTT CACAGTATA GTTAGGGCTC 1200  
AGAAATGCA TTGAGTAGC CTTATTTCT CCGTTTACA GATCTTTAA GTACACATG 1260  
CTGACTTAG CCACCCCCA GAGTTAGGA GAACTTTCC TTTTCAAGC CACTTCCAT 1320  
AAATAGGTG TTTCTGGCC TTCAAGATA TAGAACTTG CAGCACTAGT AAAGTCAG 1380

GGTGTCTGTC TCTCTACTCA ACTTATTTTG AAAATGTCTG CAGCTTCACT CCTGTAGAAA	1440
AGGAAATCTT CATATTTTGG TAAACTTTAG CCGCAATGTG CTCTGTGAG ATGTGGCAAT	1500
TCAAAATGCA GTGAATCTG CTCTCTTACT GATTCCTGAT TTTAGTGTGT GTGTGCGGGG	1560
AGTGTGTACC TATTTATAAA GAGCAAGTGT GATATGTGTG TATATGTATA TACATACATA	1620
CAATGTCCAA CACACACAAA CAAATTTTGA GAGCTTAGGA AAATCTGAAA CAGCCCTTTC	1680
ATTATCTTGC GTTACTATTC AAAATTTTCT GTCAACCCCTA ATTACAGTGT TCACATATTA	1740
GTGTGGGGCT AGGTACTTGC TTACAGGAGG AGCAATTTCC TACCAAGTGT CATTTACTTC	1800
TAAAGCACTG AGTCAGAGTG AGACCCCTGA AGGAAATTTG ACTCCAGCCC TCTCTCCAGAA	1860
TGTCTATATA GATGGGAAAC TTGGATGTCC AGCCATTTTG GTGACTGTAG AGCTTAACTA	1920
CTCAAGTTAG ACTTAGAGGC ACATAATCAG AATTCATGAC CTGTATGTGT TGGCAGGGTC	1980
TAGGAAGTCC TCTCTCCCA AGTGAATAAT ATTCTTTGCT CATTCCTGAA ATTCCACATT	2040
CATATATATG CTGTGCATTA CATGCTTCTC AATTAAGAAA TTAACTGCAT GTTTAACTGTG	2100
TGCTGTATAC ATCAATTTT TATGTTTTAA AAAATCTCAT TATGTATTTA GTTCAGCCCA	2160
GCTTTAGAGG AAAAGAGAGG CCGATATGAG AGACTTCACT CTCAATTTTA TTGCCCTTAT	2220
CGAGCACTGC TTATTAAGCC CCGTACCTTG TCCCATTTCA GAAACATATA GATCTAGGCA	2280
GTCTGTGATT CTGGAGGGCT GCGTGTATAG ATTAAGTATGT ATTAATTTGA ACTGTAGACA	2340
TACAGAAAC AGCAGAGACA GCGCCAGAGC AGGAAATTTA AAATTAAGTG AGACATTTAT	2400
GAAATCATTG GTCCAAAAA AGCCAGGGG CCGATATCTG GCTTGAATG ACTTTGAGAG	2460
GACACAGAT TTAATCTTAA TGAAGGTATA ACCTCAGAGG TCTTTCTACA GTACAGGCC	2520
AGACTGTATG AATGGAGAGA ACCAATGAAT CCAATGTGCT CTGCCATTTT TCCCTGTGAC	2580
AGTCACATTC CCGCTTTAGG AATCTGCCC TTCCAGCCCTT TACATTTAAC AAGGGAGAAC	2640
TGAATCTTTC AAGGAATTTA CAGCTTTGAG TTAATGTTTC AGTAATTCAT TTTCATTACTG	2700
TAAATTAATTT TGTAAAGAGG ATTACTGTCT ATCCAGAGAT GTTCCGACAT GGTGCCCCTG	2760
TGCATTTGGA AATCAATATA CTATTAATCTG AAATCCAAA AAAAAAAAAA AAAAAAAAAA	2820
NNAAAACTC GAGGGGGGCC CCGTACC	2847

(2) INFORMATION FOR SEQ ID NO: 101:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1394 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

GAAATGTGTC GAGGAGAGTA AATTAATCAG AGGCAAGAT TCAGTAGAGG CCAAGGGGA	60
CCCCAGAAA AAGATATGGA GCTAACTCAT CTCTTTTACA AGGGGTGGCC ATATCTTACT	120
GTTCAGAAAT ACTCAGATGA TATTTAATGT TCAATGTGTA ATTATTAATTA CGAGAGGAAA	180
GAAATTTT ACTCTGTGCT TTAATTTACT TCGTAGAAGG CTGTGGAGCA AAATGTATG	240
AATTAAGAGG GCGACTTTCT TTGTGATTTG TCGTTTCAAT GCAATTAAT TTAATTTACC	300
ATAATTTCCA AAGATTTTGG CGTTCGGCTC TCTGTCTTTT TTCTTCAATC CAGCCCTTTC	360
CTTTTTTTGG AAGGGGGGTTA TATATAGAGG TTCAATGAGG AAGTCAATG AGCGTAGAAT	420
AAAGGGCCAA GATTAAGGCGAG TTAACTTAAG AGCACTTTAT TTCTTTAGG CCTTTCTTAA	480
AAAGAAATG GGTGTGAGAT GGTGTGAATC TCCCATGATG TTGAGAGGCA CTTAAGTGGG	540
TTGAAGTATG ACATTAATAT TCCCATGTGG GAAAGAGAAA TTTCCTTTAG AAGGTGTGAAA	600
AATGCTTTTG CCGATGTGCT CATTTTGAAG CATCTTTTCC TTCTGTATTC GTTCCAGTCA	660
GGTGTGTGTC TATACAAAC TTCCCATGAG TTCTCTGCAA TATTCGCCAT TTGTAAATGA	720
TCAGTCTCTT TTCTTAACCC CTTTTCTGT TCAATCAT ATCAAGATTTG CAAGGTATG	780
ATCTATCATG CAATGTGCCC TTGTTCATCT GTGTCTTCTG CAATCTATG TCATATAGAA	840
TTCTGGGCTG CAGCAGGATA GCTGAAGTTT GGTGTGTGGA CTGAGATTTG GCGATTTAGC	900
ATCTCATGAA TTCCAGCTTC GTTCCAGCAA GCGGATCTT GCTTACGTGTC ACAGGGCCAAA	960
CGTACAGTCC TTGTGGGCTTC AATTTTCCCT CCGCTCATG AATAGAGAGG AATTACTATT	1020
TTTCTTTTGT GTTACCAAT GCTGTGACAA AAGTGTATGCT ATTATGTGTG TATTTGGGTGA	1080
TGTGTGCAAA ACTGCAAGAG CTCACTGCGT ATTAAGAGAA ATTAAGAGAGA AAGTGTAGAGA	1140
GAGGAGCAAA AAGATGAAT ATTGTGTATA GATTCAGACA TCCCAACTT TCTCTCTGCA	1200
GTCCCTGTGTC CTGATTTTTC TGTTTGTGTC AGTCTTTGT GCGACAGCCC ATTAATGCTTT	1260
GCAATGTGTC ATCTGTGAAA GGGGGTGAAT GGTGTCCAAA GTTGTGTGCA TTGTTTTTTT	1320
GCAATGCTTC TTAATTAATA AAAAAAAAAA ATGTTTTAGG TTTTATCTTA AAAAAAAAAA	1380
AAAAAAAAA ACCC	1394

(2) INFORMATION FOR SEQ ID NO: 102:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 794 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

5 GAGGAGAGG GAGTAAAGG GACTGAGCG AGCCAGTTGC GCGATTATTC TATTTCCTCT  
CCCTCTCTCC GCGCCCGTAT CTCTTTTAC CTTTCTCCA CCTTCCCTCG CTTACCATGG 60  
10 CCGAGCTCG GCGGCGACTC AGTCCCATTC CATCTCTCG TCGTCTTTTG GAGCCGAGCC  
GTCCGCGCCC GCGGCGCGCG GAGCCGAGG AGCTGCCCC GCGCTGGGA GAGAGAGCTG 120  
CAGCTCTCTCC TGTGCGGTGC AGCATCTGAT TTTCTGGAGA GATGTGAGGA AGACTGGGTT 180  
15 TGTCTTTTGA CAGCTGATC ATGCTGCTTT CCTTGGCAGC TTTCAGTCTC ATCATGTGCG 240  
GTTTCTTATC TCATCTCTGC TCTCTCTCT GTCCACATCA CTCCTAGGAT CTACAGTCC 300  
GTCACTCAAG CTGTWAGAA TTCAGAAAG GGCATCCAG TCCAAAGCTT ACCTGGAGCT 360  
20 AGACATTACT CTGTCTCTAG AGCTTTTCCA TATTATCATG ATGCTGCGCA TGGTGCAGAT 420  
CAACAGGCCC CTGAAGCTCA TTATTCCTCT CTTTCTGCTA GAGATCTGCG TTGACTCTCT 480  
25 GAAGCTGGCT GTCTTCATGT GCTGTATGAC CTATGTTGGT GCTGTTTTTA ACCGATAC 540  
CCTTCTAATT CTTCCTGAGC TGTCTATTTT CAGTGTCCCG ATTGTCTATG AGAATGACA 600  
30 GACCGAGATT GATCATATG TTGCTATCG CCGAGATCAG ACCAGATCAA TTCTTGAAA 720  
GATCCGAGC AAAA 794

(2) INFORMATION FOR SEQ ID NO: 103:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1544 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

45 TTCTGCTGCT AGTGTGAGC AAGAGTGTG TTGGGCAATT GCTGTGTGCG CCATTTCTCG 60  
ACGAGAGGCG TCTTCTTCTT CCTTCCCGCA GCGAGCGAGC TGTCTCTGCG CCGAGCTTTC 120  
50 CTGGGTGAAA AGAGTATATC CTTTCTCTCG GCGCTTACGA TACGAAAGTG ACCATATGTC 180  
GCGCAGCTCG CCTTCATGCG TGGCGCCGAG CCGAGTCTG CACTGCGCTG GATCAGCTTC 240  
TTTGAAGCTT AGCCATCTCC TGTGAGTAG GAATGAATCT GCGAGCTTTC AGGTGCTTTC 300  
55 AGCTATGACC ATCTGTGCGG TCGAGGTTACA CTCAGTCTTC CTCGCCACT CCGAGAGCTT 360  
TTAAGAGTGT TCCCTTTGCG GCGCTTCTGA GCGAGAGCAC TCGCTGAGC CTTGTGTAGA 420  
60 CTCCTACAGG GAGGAGCGAG CTGGCTCTAG GAATGGGACA CCGAGACTG GCGAGGCTTT 480

5 CAAGAGCCT GTGTGGGGG CCCAGGATC CTTAGCTGAA GCGGGGAGAC TCACCTCTCA 540  
TCTGAGAAA TTTAGCCTT TCCCTCTAGG GAGCCAGGCT TGAGGGGTAG GCCCAGAGCC 600  
TCCCTTAGGG CCTTGGGTGG GCACTGTGCG CCCTGGGGT AGGGAGGGAG ACTCAGGCCC 660  
ACACTTGGGT ATTTCATAT TCGAGAGAAA CACACACTCA GCGGCGACTC ACTGATTCCT 720  
10 ACACATGCC AGATTTTAC ACATGTGACC AGGGGCGACC AAGTCTCCCTG TCACCTTTGT 780  
GACTAGATC CTATTTCTC TATTTTCCG TGGTGGCTG GGTCTGTCTC AGCTGGGGCA 840  
GTGTGGTAA TGTTAGTTTC TGTGACACTG TTTTGTGGGG GTGGCACCTG GTTCTCCGAT 900  
15 GCGTGGGCTG GTGTGAGGCC CAGGACTGTA GTCTGGGAG CAGTAAAGCT CAGCTCTCTG 960  
TATGTAGTGA TCTATGGCT TGTCTGTCTC TTATGATCCA ATCTTTTCTT ACATCAGGCC 1020  
TTGTTTTGTT TTATGGCTAG TCTTATCTCG CTTGTTTATT TCTTGGCGG GAGGAGAGGG 1080  
TTTCTAATC TCTTCCAGC CCACCTTATT ACACCCCGAC CTGCTGGGA CTTACTGCTC 1140  
20 GCGAGGAGC AGACAGGGAG CCACAGCAG TGGTCTCTG CCGCTGTGCT GGGGTGGGG 1200  
GGAAGCTGG GGCACATGTG GCGCTTCCCT TGTGAGCAGC TCCAGTGCC AGGCTTTTGA 1260  
25 GACTTTCCA CATGATAAA GAAAGGGAG GTACAGAGT TCCAAATCCC TTTTATTTT 1320  
GCTGTTGCTT ATCTTAAT GTTTATATA TATCTAGCA TGTATCTATC AACCCAGAA 1380  
ATTTCAAAT CTCTTCAAC AATATGAGC TTTTAGCATG TTTATATTC TTCATCCTC 1440  
30 TTGTTTCCA GGTTTTGCAG GGAAGAAAG TCTGGATTA TAGATACAG TTATTTATTA 1500  
ATTGTGCTT GATTAAGAAA AAAAAAAAAA AACGCGGGG GGGG 1544

(2) INFORMATION FOR SEQ ID NO: 104:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 871 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

50 ACCAGCGCT CCGACTGTCT CACCCGGGGG CTTGGAGTG AGTTACGAGA TTCAGCCCAT 60  
TTGGCCCCGA CCGCTCTGTT CTGCGAATCC GGTGTCTGCG GATTGAGTTC CCGTTCCTA 120  
55 AGTGGGTGCG CTGTCCAGCC GCGGGGCTGG GAGTGAAGTA CCGATTTCAG CCCATTTCG 180  
CCGAGCGCT CTGTCTCTCG AATCCGGGTG CTGCGATTTG AGTCTCCGCT TCCTAACGA 240  
CTCCAGATG GAGGAGCGG GGAAGCTAGG AGGCTGATTT AGATGTGCTC ATCTACTGCT 300



CTTCTCAGGT GCGTGGGGCA TGCATATGTC GGTACAGCTC GTCTCAGGCT TTCTCTCTTT 360  
TCCGAGGCGT TCCCGGACAT ACCTTCCGAC TACGTCCAGAG CAAACTCTTC CCGTCTTACT 420  
TCCAGATCTC CATTGGAGCT GCGTTCACTA ACCTTCGATC CTTCGCTTCA CAGCATGCGT 480  
GGGCTCAGCT CAGATCTGTC GAGGCGAGCC AGGTTTACCT GCTATTCTGTC AGGCTTACCC 540  
TCCGCACTGT CAGCGCCGCC TGGCTGGAGC CCGGACAGAC AGCTGCGATG TGGGCGCTGC 600  
AAACCGTGGG AGAAGAGAGG AGGCGTGGGT GGGAGAGTAC CAGGAGGCTA ACAGGTCTCC 660  
GATCTTACAC GCGAGATGCG AGAAGAGAC CCGAGTTACA GTCTCTCTCG CCGAGATTTC 720  
TTCCGCTACG ATGGGCTGTC CTCTCTTTCG AATCTGGGCT GCGTCCGTAG CATTGGGCTC 780  
TCTCTCCGTC GCGTGGCGCT GGAATTTAGG AGCCTTACG ATGGGCGCTG CAGCTTATTA 840  
AATGCTCTTT CAGAAAAAAA AAAAAAAAAA A 871

## (2) INFORMATION FOR SEQ ID NO: 105:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

GCGAGAGTT ATACAGTGC ATTCTACTT TTTTATTAAT GCGTCATGAC TTTTTCAGT 60  
TTAGAACAA ACAGTCCAA CCGTAGGCT TCTTCCGATC AATTTTGA TCTGCTCCAA 120  
AACTGCTTTC AGTTACTAG AACTTCAGC TCCCATGCA CTGAGGCAAT TCGTTGCA 180  
AGATACAGA ATGGTTACA CATTAACT GCGAAGAT GAGAGAGTCT TAAATTTTC 240  
TTTTTAATTA GATGAGCC CCGCTTGG GACTAAATT GTCTATTC CAGAGACAG 300  
TCTAAATTTT ATTTTTTTA AAAAGAAAC TCGCCCATTA TTTTGGTGG GTTGTTTTT 360  
AATTTTAAT TTGAAAAAT TTTTGGGCT TTTTGGGCGC ATGC 404

## (2) INFORMATION FOR SEQ ID NO: 106:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

GTGACAGAG TGGAGCGGCC GGGGAGAGAG TCTGAAAGG CAGAGCTCA GAGAGAGAG 60  
GCGCTTACGA GAGAGAGAG GAGAGCGCTG GAGAGAGAG GCTGGAGAT CATTAGCCAG 120  
GATGGAGGCT GTTGTAACT TGTACAGAA GGTGATGAGG CAGCGAGATC CCGGATCCA 180  
GGGCTACGCT CTGATGGGCT CCGCTTCTCT AATGACCTCC ATTCTCTCA CTTACGTGTA 240  
CTTGTGTC TCACTTGGGC CTGCAATCAT GGTATATGCG AAGCGCTTTC AGCTCCGTCG 300  
CTTCAATGAT GTCTACACT TCTCAGTGGT GGCACATGTC CTTCAGATTC TCTATGAGTT 360  
CTTATGTGAG GCGTGGCTGA GGCATTAAC CTGGGAGTGT GACCTGTGAG ACTATTCGAA 420  
CAGCGGTGAG GGCATTAGCA TGGTGGGAGT GCGCTGAGTC TTCTCTCTCT CCGAGTTCAAT 480  
TGAAGTATG GACATGATGA TCTTTATCT CCGAAAAAAA GAGGAGAGAG TGAAGTCTCT 540  
ACATGCTTTC CAGTACCTG TCGTCCCTG GAGCTGATG TGGGGGGTAA AATTTGCCCC 600  
GGAGGAAATG GCGTCTTCC ATGCAATGAT AAGCTCTTC GTGCAATGTA TAAATGACCT 660  
GTACTTACGA TTATCTGCT TTGGCGCTGT GGCACAGCC TACCTTTGCT GAAAAAAGCA 720  
CATTGAGGCT ATTGAGCTGA TCCAGTTTGT CCGTGTCTCA CTGCAATCT CCGAGTTACTA 780  
CTTATGTCAG AGCTGTAAT ACAGTACCC AGTCATTAT CACTGATCT GATGTATGAG 840  
CAGCATCTTC TTCAATGCTGT TCTCAGACT CTGTATGAC TCTTATACCA AAGGCAAGAG 900  
GCTGGCGCGT GCACTTACG AAAATGAGC TCCAGATTT GCGAAGCTCA AAGGCAAGTC 960  
AGAAGCATG CTTAGATGAG CCGGACCTTA AGTCCCTGAG GACTGAGCT TTAGGCGATG 1020  
TCCGTCAATG CCGCTTCCAC CTACAGCTGT GACCAAGCT TATGTGATCA GACTGAGCA 1080  
GGGAGTACG CTTCCCGTCC CAGAGCTGC TCTACAGGGA CCAAGGCTTT GATTCTCAC 1140  
CGATTCGCCC GGGGAGAGTC CAGGATGTC GGTCAATTC TGTCTGAC TCCAGAGCTG 1200  
GGGCTTAAAA GCGCTGTACA GTTATTTCCC CTTCCCTGAC TTAAAACTTG GAGAGGAGC 1260  
ACTGAGGAGT GCGGCGAGAA AGGATGTCGT GCGCTTTTC GTACAGAGAA AAGATGAGC 1320  
AATATGTC CTTGAGAGCC AGTCTGATC GTGACCCCA GACTGAGAG CAGTTAGTTC 1380  
TGGGCAAAAG GTCAAGGTG GCGGGGAGCT GGGATACAG CCGTGGAGAG CTGCTTACTC 1440  
AAGTGTGTC TTAAATTAAG GTACAGAG GAAACAAAA AAAAAAAAAA AAAAACTCA 1500  
GGGCGGCGCG TACCGAATTC GCGGTATGTA TCTGAAAG TC 1542

## (2) INFORMATION FOR SEQ ID NO: 107:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2127 base pairs  
(B) TYPE: nucleic acid

## (2) INFORMATION FOR SEQ ID NO: 107:

329

330

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

5 GGTAGCTCAN TCGAGTGAAG TATCTCTACT GGAAGCAAG CCTTTTCA AGATTATTA 60  
ACTCTTCCT TTTCTTTTG GAGAGTCTT TTTGTTCTGA TCGACCAAT TCACTGCAG 120  
10 AAGCAACAA GTATTCTTAG CAGAAGATG GAGCTTGAG CCAATTTGCG GAGGGCCAGT 180  
ACATTATCT GAGCTCTGA GTGTGAGAA TATGAGCTCC ACTCTCACT ATATTCAAR 240  
CGATTCAGAC TTGAGCAACA ATAGCAGTTT TACCCCTGAT GAGCAAGGA GACTTAAGT 300  
15 AAGAGATGTT GTACCTCAGG CGTTCTTAGA TCAATATTTA TCTATGACTG ACCCTTCTCG 360  
TCGACAGAGG GTTGACACTG AAATTGCTAA GCACTGTGCA TATAGGCTCC CTGTGTGCG 420  
20 CTTGACACTC GGAAGACAGA ATTGCGACTG CTTGAGAGAG ACCATATGGA CTTTGGGCTC 480  
AGACATGAG TGAAGATTC GAGGAACTC TAGCATCTC CATCCAGCG CTTCAGTTA 540  
TTCTTGAGA TCATTGACA GCTGAGATC TGGTTCCAT TTTTAATGA TTTTAAAG 600  
25 ACCTGATGA AGTCAGATA GGTGTTCTTA AACACTTGA TGAATTTCTG AGCTTCTTC 660  
ATATTGACA AAGAAGAA TATCTTATC AACTTCAGA GTTTTGTG ACAGTATA 720  
30 GTAGAAATG GCGGTTTGA GCTGACTG CTAAGAGCT GATTTTACTI CTAGAGTAT 780  
ATAGTCCCG AGATGTTTAT GACTATTAC GTCCATTCG TCTGATCTG TGTGACGA 840  
AAGTTTCTC TGTGCTTGG ATTCTTACA AGTTGCTAG CGAGATGCG AAGAAGCTC 900  
35 ACCGCGCAC ACCACACAG TTGAGATG ACCGATCAA TGAAGTTG GAGAACTTG 960  
CGAGATGTC CAGTGTCT GTTGGCAG CTTTGTCTT TGTCTGCCAG ACTGTCAATG 1020  
40 AGGATGACTG CTTTCCCATG GACCAATTTG CTGTGCACT CATGCCCAT CTGTAACCT 1080  
TAGCAATGA CAGGTTTCT AACGTGAG TGTGCTTCC AAGACATTA AGCAAACTC 1140  
45 TACTAGAAA AGACTATTC TTGCGCTCTG CCAGCTGCA CAGAGGCT GTGAGCAGA 1200  
CCATCATGCG TCTTCAGATG GACCTGACA GCGATGTCAA GTATTTTGA AGCATCCAC 1260  
CTGCCAGTAC CAATAATCC GAAGATGCA TGAAGACAG GTCTCAAC TACTAGAGG 1320  
50 CTTGAACTC GGTGTCTTC CTGCTTCAT TGAAGCAG GTTCAATG CATTCGCAC 1380  
CGATGTGACC TGGATGACT TTGCGGAG GAGAGACTT CTCTCTCC GCACTTCAT 1440  
GAGGTGGA GTTGCTTACA CCAATACA GGAATTTCAA GAGTCAGAG AAGTACAT 1500  
55 AAGACTATT ATCTATCTT GACTTTAG KNNRRNNN KCTGAGNSA TTTATTTSH 1560  
CNRARGSH WYNAWSCJK SKECTVWCC KSRSTGRMG MRRCTCTACA ATTTGRLAK 1620  
60 CNYTKSCT RMWGAARKS GGCASGACC AGAGACTCC ATTGCTTTCT CTGCTTTTA 1680

TTTAACATC GACAAATGA ATTCTTACG CCGAAGGCA GAGCTGTCC CAGATGTGA 1740  
5 AGAGCTTC AGTATCAGC CTAACTCTT TCTCCAGGA AGGACTTCT GGGCTCTGTG 1800  
GCCAGCTGC CAGCCAGCC CTGTGTGCA ATCTTTTGT AGGTGTGCA ATGGGAAGG 1860  
AGGGTTTTT ACATCTCTA AGGACCTCA TCCAAACA AGTAGAGTG ACTTAACCTC 1920  
10 TTAAGCGAG CATATGCTG TAAACATTA CAGAAATGTT CTTGAGTCT TGTCTGTAT 1980  
TTTTTCATC TGTCTATAC CTGAAGGAA TTTTATGAC GTATATGTA TGTCTGTGT 2040  
TTTTTACTG ATCATGATCA GCTCTGAGT GCACTTCTT CACATCTCT ACATCTCT 2100  
15 GACCACTCT GCGAGTCTG CAGCTTTAA TCACTCTGT TAACTGTG TGGCACAAT 2160  
TCTCTGTCC AATTAATAT TATTAATAG ATCTATGAG AGAGATAT ATACTTTTGA 2220  
20 TTGTTTTCTA GATCTTACC AATTAATGA ATTGTGACC TGTAAATAA AAAAAAAAA 2280  
ACTGAGGG GCGCCGCTAC CCAATGCC GATATGATC AATCATC 2327

(2) INFORMATION FOR SEQ ID NO: 108:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1062 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

60 GCGCCGCGAG GCGCAACAG CTTTCTGTA GCTCTGGTC CAACCGACT AGCAAAATC 60  
120 TTCTCATCC TCATCATCT CTCTCTCAT CCGATCTGG TCCAGSTCC TCTCCCTCCC 120  
180 ACACAGAGG TGGCAGAGT CCGACTGTG TTCTCTGGA CTTTCTGAA GATCTCTTC 180  
240 CTCTTCTG TCATCATCT CTCTCTCTC TTCTCTATC TCATCATCC TCATCATCCA GTTCTGAG 240  
300 CCGCTCAGA ATCCCATCC CCGCCGCGA GAAATGACA GGAAGCGCG GTACAGCTCT 300  
360 TATCTTTCAC ATGACCATTA CCAAGGCCA AGAGCTCTAC AAAGAGAGC TGCATTAAGA 360  
420 GAAGAGGG TGGCTCTCAT TGAAGATA CTTGCGGCA TGAATGATC AGAGCTGAA 420  
480 CAGAGTTCT CCGTTTTTG AAGATGAG GATGACCA TCACTTCC TGTCCAAGG 480  
540 GACACTGAC GCTTGTGAC TTATGCTAT CCGTAGAGG CATTTGAGC CATTTGAGT 540  
600 GCGCAAGC TCGCGAGCC AGATGAGAG CCGTTTATC TCTCTTTG GCGCGAGG 600  
660 SMTNCTGA AGAGAGCTA TTCTGATCT GACTCCACC GGAAGACTT TGAACCA 660  
720 CCGTAAAGA GCAATTTGA TTTCTTGC TTTGACAT TGTGAACA GCGCCAGAG 720



(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

5 AGCATGAGCG CGATGGCGCT GGTGGCGAGT AGCTGCTGG GCTTGCTGCA AAGCATGGT 60  
GCTTTTGCG CGATCTGGT GGTGGCTAC TACGTATTG GCATATTGG GATCACTTG 120  
TTTAGAGGG TCATTGAGC TCTTCTGGA AAGAGAGCC TGCGCCCTGC CATAGCTCG 180  
GCGCCCTGTC CGAGCTTCCA CGAGCTGGAG TACTGGGCA ACACTTCCA TGACTTTGG 240  
GCTGCGCTGG TCATCTCTGG GAATTCATG GTGGTGAACA ACTGGAGT GTTCTCGAT 300  
GCATATCGGC GCTACTGAGG CCGTGGTGC AAGATCTATT TTGTATTGG GTGGCTGTC 360  
TGTCTGTCA TCTGGGTCAA CTTGTTCTG GCGCTGATC TGGAGAGTT CTTTCAAG 420  
TGGAGCCCC GCAGCCACCT GCAGCCCTT GCTGGAGCC CAGAGGCCAC CTACAGATG 480  
ACTGTGAGC TCTGTGTCAG GCATATTCTG GAGGAGCGCG GCGAGAGTGA CTTCAAGAG 540  
AGCTGAGCC AGCAACCCCA CTTGTGCTG TCGAGGTGAC GTCCGGGCTG CCAATCCAGC 600  
AGCGCGGCCA GAGAGAGAG GCTGGCTTAA CACAGGTGCC CATCTGGA GAGCGGCCA 660  
TGTGTGGCC AGCAGAGCAG GAAGAGACCT TTCTCTGAC GAGCCACTTAA GCTGGGACA 720  
GGAACCAAGT CTTTTCGTC TGCGCCAGCA ACCATCTACA GAACAGTGC TGTGCTTCA 780  
GGAAGCGCC GTGCGCTCG CTTTCTTTTA TAGTGTCTC AGTGAGANT CCGTGTGCA 840  
CTCCAGAGG ACTTTTCAGA CAATAATGCA AGAAGCAGG GCTTCGCCG TGCCCTCAG 900  
CTTTCGCTGT CCGTTGCTG CCGCGAGCC TTGGGAGCCA CAGCGCTGAC CAGGGCTGC 960  
ACAGGTTAAC GGTGAGTCTG TCTCATCTAT TCACAGCTGG GAATGATCT ATATCTCGG 1020  
ATTTTAGCCC AGCAGCAGAG GGTACGTTTC AGTTTTCCTC TCTTTCCATA GCTGTAGGC 1080  
CCTTCTGCG ANTGGTCTC ATCTCTCTTA ATCTATTAT GGGTCAGTT TCTGTCATGT 1140  
CCCCAGCTC CCATCACTGC CAGCCACTCC CCACAGAGT GCGCTGCTCA TCGAGCTGG 1200  
GCTTTGACT CCACACTGTC TACCGCTCT GTGTGGAGCG CCGTGTGCA AAGCTTCAG 1260  
CAACAGCTT TCCAAATGGA AGTTGTACT GTACGGCTT TACATCAGC AAGAGCAAA 1320  
TCTATAGCT GCTGAGGCTC GTGCGCTCAT AGATGCAAT AATATGTAA GTACTTAAA 1380  
ACAGCATAG AAGAAAGCTA ATGCTTTATT CTCAAATATG ATGTCTACAT AGAAGAGCCA 1440  
AATATTTAA GATATGTAG AATTCACCA GCAGTTTGG AGCGCGAGCG GGGTGGATCA 1500  
55 TGAAGTCAG AGATCGAGAC CATCTGCTT AAGAGGTGTA AACCGCTCT CTACTTAAA 1560  
TACAAAAT TGCGCGGCG CAGTGGCGG GCGCTGTGCT CCGAGTACT GCGGAGGCTG 1620  
AGCAGAGCA ATGCGGTGA CCGCGAGC GAGCTTCCA GTGAGCGAG ATTGCGCAC 1680

TCCAGTCCGC AGTCCAGCCT GGGCGACAGA GCGAGACTCC GTCTCAAAA AAAAAAAAAA 1740

AAAAAAAAA A 1751

(2) INFORMATION FOR SEQ ID NO: 111:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1117 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

AAATGTTGG TCGTACGATT TGGTTAATT CTTATTATG AGTCTCTGG AGAGCAATGT 60  
CGATTAACCTA ATCCCAACCA ACNTGTCTT TTTRATGTTG TAGTGACAG CAGAGATTT 120  
CAAGAGCCT TCGTAATATC TGTAAAGAGG CAGCTACAGC AGGCATCATT GCGTGGTGT 180  
ATGGGGGAT ACCAGCTTTT ATTCTGCTA AACAACATA CATGAGCAG AGCCAGCGAG 240  
AAATTATCA TAACTGGTTT GATCTGTGC ATCTGCGCA TGTGTCTGC ACAGAGGCT 300  
TCATCTGTTA TCGCTGCGC TGGGTTGGA GAATGCACT GTTGTGACT ATATTCAACA 360  
CAGTACAGC TAGTCTGAT GTATACGGA ATAAAGTGC CTTAAGCCAT TTGTAAATG 420  
CAGAGCTGT CAGCGAAGT CTTTTTNGA TAAAGTAGG CCGTGTGTC CTGTGTGCTG 480  
GTGCGATAT TCGAGGCTTG CTGGCACTC CTGTAGAGG CCGTGTGATG GCATTTTCA 540  
AGTACTCTG TCGAGCTGT CAGAAAGAA AACAAGAGA TCGAAGGCA CTCATGAGC 600  
TAAAGTGA AGAGTGGAA GCGAGACTAC AAGTTACTGA GCACCTGCTT GAGAAATTTG 660  
AAGTATGTT ACAGGAGAT GAACCTGGA ATGATCTTAA GAATTTGAA GCACCTGCTTAA 720  
ACCTTCTTAC AAGCTTCTA GTATATGATA AACAAGACA GGAAGTGAAG TCGTCTGAC 780  
TTGAAACTCA CTGAGAGCT GAGGGAGCT GCGATGTCG ATGATGCGA ACGAGCAGGC 840  
CACTCTTGG TCGAGCTGCT GACAAATTTA AGTCTGTGTA CCGTGTGTC CAGTGGCTTG 900  
CTCTGTCTT TTCTTTTCT TTTTACTTAA GAATGGGCT GTTGTACTCT CACTTTACTT 960  
ATCTTAAAT TTAATATAT ACTTATGTTT GTATTATCTT ATCATATAT GCAATACATA 1020  
ATATATCCAC CCACCTAGAT TTTAAGCAGT AATAAACA TTTCGCAAA GATTAAAGTT 1080  
GATTTTACA GTTAAAAAA AAAAAAAAAA 1117

(2) INFORMATION FOR SEQ ID NO: 112:

60

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1313 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

55  
50  
45  
40  
35  
30  
25  
20  
15  
10  
5

GGCAGAGCTT TTCTTATATT TTAAAGTGGC TAGCAGATA TTATATCTTG  
TTTGAGACTA CCAACATAC TAGATATGA AGATCTTCA GAGAGATAT ATTGCTTTA  
ATGTGAATA ATTTCACA ATGTGCTTA CTTATATATA GTATATTAAT TGTAGATAT  
TGTATTAAT AGTTGTAC CTTTTCAT TTATATATA GTATATGAT GTTACTTTT  
TTTGAATA TGCAGAGAT CTCTTGAAC ATGATATGG GCTCTCAAT ATGCTGTGAT  
TACTAGAAA AGAAGAAA GCAATCTTC TTATATATCT TGGCCATAT TGTAAAGCT  
TGTCTTCAAG AACGATATC GCAATTCGA AACGCTTGT TAAATATGAG AAGGCGACA  
TTAAAGAGC TTACAGAGT GCTCCATAG TAGACATGA ATTACTTGA TTGATCTTC  
GCTATATTA GCGAGAGCT ACTTCCATG CTCAGACA TGAAGAGCT GAGATATTA  
AAGTTTACA GTCCAGATC GACAGAGAC TGTGTGTGC TTAAAGCTGA AGTTACATA  
TGTATTTT GGTACTTCT TCTCTCAGA GTGCATATC TTTCGAAA TTCTTGGTT  
TGCAGAGCT TAGTACAAA GCGAAGAA ATTATACAG CATCTAAA GAGTGAGAA  
TTTATATCTT TAGAGAGCT AGTTTGGCC AACTTAGAT TTACGTTTA TTATACATA  
GTATTTACA CTCATCAAA ATATATGTA AACATCTGA TTATATAT TTATCTGCC  
CTTTGTATA AACTGAAT TTGTGAAT GATTGTGCT GCTCTCAG CTAAGATA  
TTTTGTGA ATGAGAGCT GATTTATGT GTGATATC CAGAGAGC CAGCAATTT  
ATTCAAAAC AATGTATCA TCAATGAT TGTCTCAT GTCCAGATC ATGTATCA  
GACAGCTGA AAGGAATCC TGTCTAGCT GCGTGTAT GTTATAGC GAAATTCAG  
AGAAAGAGC CAATATGTT TTTCCTTT GAAAGTTT TAAATATAT TTCATGGTC  
TTTTTTTA TTATATATG TGCATGTTA CATATATGT TGGATGTCT TTGACCTTA  
AATGCTTTT TTTATATG AGATATGTA CTTATTTAT TTATATTA TGTATCTCC  
CTTTTAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAA 1313

(2) INFORMATION FOR SEQ ID NO: 113:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1654 base pairs  
(B) TYPE: nucleic acid

- (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

55  
50  
45  
40  
35  
30  
25  
20  
15  
10  
5

AAGAGAGAG AATCTTCT TTCTTCTCT CAGTACAG AAGCTTCT CTACATTTG  
CTCTACAT TTATTTTAA AGTATCTT TTCTATGAT ATTATATCT GCGATAGCA  
TGTATACAG AACATCTTT CATACAGAC TGAAGGAG CATCTCTTA CTTAAATAG  
TTCTGATCT TTCCCTCTA TTATAGAA ATTATACA TTCAAAAT AACTTAATTT  
TTCTTGAAT TTATTTTAC CAGTACAA ATCTTTTT GATTTTGA GATTTGCT  
CATAGATCA CAATATCTA CATTCATTA GATGATATT TTCTATGAT CCGCAAGAG  
CATGTGAT CCAATGGCT GTTGAGCTA GAAATTTCC TGGCCCTGT GACCTGTAA  
GCTCTCTCT CCGAGCTGT TAGATGGTG AAGAAATGA GAGATGCA GATGAGAG  
AGATATCAT GACTGTCT GCTCTCTG GTTCTGTCT TGGCCAGAG TTATGAAA  
GCAAGGAAA TAGATGAT TCACATCTC AGCTTTCGA GCTTTTGT TTATTTTGT  
TTCTTGAAT TCCCTGTGT AAAAGGCGA AGAAATGTA CTCATATCT CTACAGACC  
ATGCAAGCT AACGAGTAG TTATATGAT GATTTTGG CATAGATAT GGTATATC  
AATCTGAGC GACATTTCC TTCAATATC GTCCAGCTG AGTATCTGT TCTCTCTCG  
CCCATTTCT TCACCTCT GCTTACCT TGGCGAGAG AGATGTAT TTCTTTTGT  
TCTCTCTCG TTATGAAA AGCATCTT AATATATCT TTCAAGCT TTGGGTTGT  
TTTGTATTT TTTTCTCT CCAAGAGCT CCGTGTGT ATTGATTTT GTATTTTAT  
ACATATAT GATTTTATA AGCTTACA GATTTTATA TTATGTGAA AGAATGAAA  
GATGAGAG AAAATATTT AATATCAGC TCAGTACA AGTGTCTAG ATTATTCAT  
TGGAGTCT CCTTTGTGA GATTTTGA ACACCTAG ACCTAACTG TGTACAGAC  
TTCTGATGT TTAGCAGTG CTAGTATTT CCTGTATAG ATTGTATAT TACTTCTTA  
TTCTTATCT CTCTTCTCT TCAGATTA TAGATTTAA AATGAGTG GATATATCA  
AAAGGTAT GTATATAT AATATCTTA GTGAGATGA AATGTATTA TATATCTCA  
TTCAAAATTA TGCAGTTAG TAAATCTCA GGTATATCA AATATCTTA ATATGCTGT  
GATCTACT TGTCTCTTG CTAGAAAA TTATAGAG GATCTTGA TTGTGGAAGC  
CAATATGTA ATATCTATG TTCTAAAT TGGCTATAG ATAAATAT AAGAAATAG  
GATTTTAT CCGAGATAT GGTGTAT TTATGATAT ACCGAGAG ATGTATGAG  
TAAATCAG TTGTAAATA TGTAAATAG TCATATATA ACAAATCT GATTTATTC  
CAAAAAAAAA AAAAAAAAA ATTCAGGG GGGC 1654

5 (2) INFORMATION FOR SEQ ID NO: 114:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1171 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```
15 GGCMAACTTT CCCCMAAGC TTGMAACTT GCAGCCGAA ACCTTGATC GTTAAAGTT 60
GGTTTCGNC GCGCCCTCG CCGAGAGAG CGCAATTGC GTTCCGCGAA GTTGGCCCT 120
CAAGCGCTCG GCAGCCAGCC ATGTCTCGA CCGAGAGAG CGCCCTCGG CTACAGGAC 180
CTGACCTCA TCTTCTCGG CCGACTCGG CCGGATAGG GMAATTTC GACTGTGAG 240
GAGTGTGTC TGGACTGCT GTTGACTTC TTACCCGAG GGTGACAA AGAGAAGTC 300
25 ACACCACTCA CGTCAAGCA ACCTTATGT CAGAAATGG TTAAAGTGT CAATGACTCT 360
GACCGATGGA GTCTTATAT CCTGTAAAC AACAGTGGCA AAATGTGGA ACTGAATTT 420
GTGGATTCC TCCGAGGCA GTTTGATTC AGTGTAGTT GTTTCAAT CAANTTAGC 480
TCTCTTCTGC TCTTTATGA ATGTTGAGG AACCAATGA CTGAGACAT TCACCCGCA 540
ATAATCGGG AGAGCTCTA TGGGATTT CAGGAAGCT TTGATCACT TTGTACAGG 600
ATCAATTGCA CCAGGAACC AGAGGAATC CGAGGGGAG GCTCTTAA GTACTGCAC 660
35 CTCTTGTGTA GGGCTTATG GCGGCTCT GATGAATCA AGACCTTCA AAGTATATG 720
TGTTCAGGT TTTCATGCA CTCTCAGAC ATTGAGAGC AGCAGAGAA ACTGAGTCC 780
TATTTCGAGA ACCACTTGT GGAATTTGA AGACCCGAG TATGATATC TCATGACCT 840
TCATGAGTG GTAAATGGA GCAGAGTGT CTGATGCGA CATGAAGAA GACGACTTT 900
45 AAACCTTATC ACCATGCTG CTATCCGGT GTTACGTAC CAATATGTA TTCTTATGT 960
GGTAATGTC ACTTGTATT ACCAGCCAG CCGCTATGTA CGAGTGCCA ACTTAGCAA 1020
TTACTACATT GCACAGTTC ACCAGTATT CAGGTGGCAG CACAGACT ACTCACTTG 1080
50 GCTACCTCG ANTTAGATT CATTTAAAA TGTCTGTGG GAAAGCAAT TCAGACAGA 1140
CAGGAGAGAA, AAAAAAAAA A 1171
```

(2) INFORMATION FOR SEQ ID NO: 115:

(1) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 842 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

```
5 GGTCTGCGC GGAATGCAT GAGCTGCCA TGTGCTCTT AGTGAATCG GTTTCGGTCG 60
10 CTCTCCGCTG TTTCGCGGCG TGGATATTG CTTCCGCCA TGGGCCGCA GGGCAAAATG 120
GGCACGAGAG CGAGAGAGCA GATATTGAA GAGACAGAG AGACTCTGAA GTTCTACTCG 180
CGATCATAC TGGGGGCCA TCCCATTTAC TGCCTTGTA CGTTGCTTT GTTTTACTCA 240
15 TCTGCTCAT TTGCGGCTG GTTGGCCCTG GCTTTATGTC TGCGAGTGTG TGGGGCCAGC 300
TACCACTCTA TGAGCTGAT GCGAGAGCA GGTTCCTCG AGCATGGGC CTGATGAT 360
GTTGGCATGG ACCTACAT GCAGAGGCG ATGGCAGAGC ACCTTAAGCA TGTGATCTTA 420
CTGACAGCA TGTGTCAGT GCTCAGCTC TTCTCTCTCT ATGCTGTGTC CTTCCTGCTT 480
CTGCTTCAG GCGCGGCTT TTACTCTCG TGGGTGATG TCTGGGCGC CTGCTTCACT 540
GCAGACAGT GCACCCGAG ACCAGAGCAG ATGAGAAAC GGCAGGCGC ACAGAGCGG 600
CGGAGATGA AGCGTTATA GCAATGACA TTGTGGGCA AGGCCACTG CCGTGGGTG 660
30 CTCTGTGAG GTGCGAGCC CTTCTGCTT CGAGCAATGA GGTCTAGTC CAGGGGCCA 720
AAGCAATCTG AGTATTGGG TATACTATA CTCTATAGG TCGTTGAATA AATGCTTAG 780
AATGTGAAA AAAAAAAAA AAAAACTCG AGGGGGGCGC GGTACCCAT TTCTCTTAA 840
AT 842
```

(2) INFORMATION FOR SEQ ID NO: 116:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1640 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

```
50 GGCACAGGCG GCGGCGAGCG GTGGGCGCG CGCCCCCGCG CCGAGCGCT TCCCTTTCC 60
GTCCGGAGCG GCGGGTCGCG GCGCCAGGCG ACCCGGGCG ACCGAGAGCG GCGAGAGAT 120
GGATTGCGCG GCGCTCCCG CCGGATGAA GAAGAGGAA GTGATCCGAA AATCTGGCT 180
AAGTGTGCGC AAGAGCGATG TCTACTACTT CAGTCCAAAT GGTAAAGAT TCAGAGCAA 240
GCGCTAGTTG CGAGGTACC TGGGAATAC TGTGATCTC AGCATTTTG ACTTCAGAC 300
```

TCGAAAGAG AGTCCATGTA ATTACAGAA GAGGAAAG AGACTGCGA AGGAGCTCT 360  
GATGCAAAAT AAGGATTAAC GAGACTTGA ATACAGACTT GCGATTGA GAAAGAGCT 420  
CAATTTCGA AGAGCCGTA ACCGAAATC ACAGATATC CTAGATGTA AGTGAAATCA 480  
GAGCGAAG GATGAAATG AGAGCAAGT GAGCTTTCT GCGAGAGAG GCTACAGAGA 540  
CTTTAGTCA TCGAGATGA CAGAGCAAT TATTAAGAAC ATGAGATAC CCGAAGCTT 600  
TCAGAGATT GATCCAGTAG CATGATGAG AGCTTTAT CTGCTATTC GAGTCTTTG 660  
GAGAGAGT CTGCGCAAT CAGAGGCGA GTCTCCCTG CTGAGAAA GAGCTGCTG 720  
TTTGACTTA GAGATCGA CCCCTCTGA AGCTTTAT TGTACAGAT GAGAGTCAG 780  
GAAAGAGAG AGCGATGCA GCAATGAGC AAGAAATGG AAGAGAGCT GATGCGAGAC 840  
ATCTGTGC GAGTGTGTA TACAGAGAG ATGATATGG AATGAGAG TCGAGATGA 900  
GCGTAAGAT ATGATCAAGT AGCTTGAC GAGCTTCCC GAGAGAGAA TTGCTAGAA 960  
ATTGAGAAA AATGTTTCA CTGCTTTG CCGTAAAG AAAAAATGA CCGAGAGCA 1020  
TGAAGCTTT TATAGCACT AACCAATCC TTTTAAAG TATTTTAT GTATATATCT 1080  
ATTATCGAA AATCAATGT TATTTGAT CCTAGAGCT AAAATATGC TTTTGTAAAT 1140  
TCAGAGAGA CCGTAAGAG AGCTGAGCT TTGATGCC GATGCAATCT ACTGAGATG 1200  
TAGCACTAC GTAAAGAT TGTTCCTCC AGAGTTTAA TAAAGAGAA TCGAGATTC 1260  
TAAATAAAT TCCAGTTAA AGATTATGT GACTTGACT TATTAAGCA TATTTTATA 1320  
CTTTATGAA AGGAGAGC TGTACATGT TCGATGCTA CTGTAAAGC AATTAATGA 1380  
TTATTTCCA CAGAAAAAA AAAAAAAAT MASTGABR GSKGGRBR ATGABRACC 1440  
CGATATGAS MTCSTATKA YTTACATCA ACTGTGATCC CCGGCGCTTA GCTTTGAGAT 1500  
GCGAGGTGAG AGAGAGTAG CCGAATAT TCGAGATGA ACTATGCTT CTGAGAGCT 1560  
GTGAGAGAT GTGCTTTC GAGATTTCT AAGATTTCT GCGTAAAT TCACTGAGC 1620  
TGTGTAAAT TTTTTCCT 1640

50 (2) INFORMATION FOR SEQ ID NO: 117:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 952 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

60 TGAATTGAG AAGCACTTG GAAAGCTAT AACCTATCA GAAAGTCTT TTAGCGAGC 60

TCGTAGCTT CTAGATGAT AACAAAAAA TGAATATAT TGTGAAAT TAAAGCTTT 120  
ATTTAATTT GCAATTTTT TCAATGTTT AGATATCTT AAATTTTAA TTGTGAAATC 180  
ATTTCCTGC GAGTAACTT TATGAAAT ATTGAGCTCA TGAAGCTGA AGTATGTAG 240  
CTTTTGTGG AGCTTATGG GACTGTGAG AGATATGAT TGTATCTTGT ATCTTAAT 300  
CTAAAGTAG TTTTAAATA CTCCAAAG AATATGATC TCCTTCTGA TGTAAAACT 360  
CATCTTGGG GTAAAGATT AATGTCCA AGGTGTGAC AGTTGATAG GTCAAGAGGA 420  
GATAGCTGG GAGTGAAGT CTGCGATTC AGAGTCAGC GATTCAGCA GAGTGTAT 480  
TAAATCTCC AGTAAAGAT GATGATGAG GAGAGGGGTA GAGTGGGTT ATTAAATAC 540  
AGCTGCTGT ATTAACTAT GATGTGAGG GAGAGAGAC CTGAGAGAA GAAATGACT 600  
ATTCCTTTT TTGAAAGAG AAAAAAAT ATTTTTGT CAGTAAAT GATAGAGAT 660  
TCGATGTCC CTAGAGCAA GAGACAGTT CGAGTAGAA GTAGAGCTG GAACTGAAA 720  
ATTGAGAAA CATGAGGGA AAGAAAAAT GCTTTCTCT TATTTGCG AGTTGCAAT 780  
GAGGAGGAGG GACTGTGTT TTGTGGAAT GTATATGAT GTATATGAC ATATATGAC 840  
CGAGTCTGC TGAATATA AGGTCCAAA AATATGTA AATCTGCTT TTTGTAAAT 900  
TATCTAATA AAGCCCACT GAGCTCCAA AAAAAAAGA NN 952

35 (2) INFORMATION FOR SEQ ID NO: 118:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1256 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

45 GAGCTATAG GTAAAGAGC TGTATATCC TCGAGAGGCG CCGAGAGAGC TTGCTGAGTA 60  
CGAGCTTGG CTAGAGCAG AGAGCTGCG TGTACATAT AGCTGTGCTT GAGAGAGCT 120  
GCGCGCGCG GCGCGAGAG TGAATATGT GCTTTGAGG GCGAGAGAGC TGTTCATTC 180  
ATCTTTTAC CCGAGATAT TACAGATGT GCGATATAT AGGATGCTA GAAATGTGT 240  
TTGATAGAAC GATGCTGAG ACAGAGAGC GAGCTTAAT CTTTCTGAG CCTTGCTCT 300  
ATGAGAGAC TCTCTCTGAC TCCACTGTC ACTCTTTCAG AGCAAGATC CTCTAGAGAA 360  
CGTAGAGGG AAGAGCAAT GCGCAAGAC ATCTGAGGTG AAGAGAGGCT AAGCTTGAT 420  
GAGTGAACA AGCTAGAGGT GTTGAGATCA GAGTTTACC AGCAAGAT AGACTGAG 480

540 GAAGAGTCCA AAGACTTTGT GGCAGAAATT GGCAGTTTC AGAATAATGT TGTGTGTTTA 540  
600 ATTGAGCTTG TTGATCAACT TCGAAGAA GCGAATAATG AAGAGATGAA GGCATCGGT 600  
5 GCTCGCAACT TGTCTAATTC TATAGCAAG CAGAGAGAG CTCACAGCA GCACTTCAA 660  
720 GGCCTAATAG CAGAAAGAA AATCAGCTA GAAAGTATC GGGTTGAATA TGAAGCTTTG 720  
780 TGTAAAGTAG AAGCAGACA AATGAAATTT ATTGACCAAT TTATTTTCCA GAATGAACT 780  
840 GAAATTTTGG CTTTTHATGT AGGAGGCAA NACAANAAA AGCCTCTCAA ACCAANAAA 840  
900 ACCTCTGTAG CATTCCAGCG GCTTGACCAA TCACTATATGT CACAGAGGT GCGGTGTAG 900  
15 GAAATCAGCC CCGTGAGAC AGCACTACAA GTCTGGGGGA GCGAGTTTGA ACATCAGTGC 960  
1020 ACAGCTGCTG CTGTGCGCCC TCGAGTGTAC GTTCTCACTT CTTATGCTTA GTTGGAACTA 1020  
1080 AGCAGTTTGT AACTTTTCAT CTTTTTTTTT GTAAATTCAC AAGCTTTTGG AAGGAGAAC 1080  
1140 AATPAATTTT TGTTTTCAA TGGCTTGATG TACTTTTTTT CTGTGCTCTT TTGAATATGT 1140  
1200 TTTAAGCTCT CAGCAGAA CCGTGATTC TCTATCCCT ATTCACAAA ACAAAACAGG 1200  
25 CAGTGTGAC CAGTACCTT TAAATTCGAT CACACAGTG AUTCAGACAG TACCCAC 1256

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1143 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

30 (2) INFORMATION FOR SEQ ID NO: 119:

35 (A) LENGTH: 1143 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

40 GCGCCTACCA GCGCGGCTGG TCTGCTGCG AGCCGGGCGC CCGGAGTGGG GCGCGCGCAT 60  
GTACTTTCCA CATTGAGTAT TCGAAGAA GTGATCTGAA CTCGACCAAT TCTTTATGGA 120  
TACATTAAGT CAATTAATAG AGCTGACTA CTTGACACAC TGGCTGAGC AATCATGAC 180  
GTTGGAGTGG CCGCAGTGA ATGAAATCCA AATACCGGTG TCATGACAG CCGGGGTATG 240  
TGGCTGACAT ATGCAATGGG AGTTGGCTTG CTTCAATATG TCTTACTAG CATTCCTTC 300  
50 TTCAAGTATC CTGTGCTGTT GACTTTACCA AATATTTATC ATATTTATAC ATATCTGG GATGTAGCTA 360  
TTTTTGCATG CAGTGAAGGG AACACTTTTC GAAACTCTTG ACCAGGTAA AGCAAGGCTC 420  
CTAACTCATT GCGAACAACAT GCACTATGGA GTACAGTTTA CATCTCAAG GAGTTTTC 480  
55 ACAATTTCT CAATATTTCT AATTTTCTG GCAATTTCT ATAGAGCTA TGAATCAACT 540  
CACTTCATCC TAAACACAG TTCTCTCTG AGTGTACTAA TTCCAAATAT GGCACAACTA 600  
60 CAGTGTGTTT GCACTCTTGG AATTAATAG TATTAATATG TTTTGAATCT GAAATAAAT 660

720 TTTACAGCTA CTGAATTTCT TATTAAGGAG GAGTGGTTAG TAACTGCAC TGTTCCTSTG 720  
780 ATATGTGAA ATGAGAGCTA TTTACATTTG AGGCGCAATG GCTGTGCTCT CAAGTCTGT 780  
840 TTTGAAGTGC AGATTTCCAT TAAATGATGC CTCGTGTTAA TACACCTGGT ACATTTCTGA 840  
900 AGAGGCGTT TATTAAGCAG CTGGCAGGC CCACTTATA AGTTAAGGG CATCACAGT 900  
960 AGGCTGATGT AGATTAATTC AAGGAATTA GAGATTTGTA AGAACTAGG ACCGCTTAA 960  
1020 CTTATATGA ATGGCAATG TTTTAAGAAA AGAACATTC CAGTCAITCA GCTGTGTTA 1020  
1080 TTTAAGCAG ACTTACATGT AATCCGAAAT CCTCTCTATA CAAGTTTAT AAGATTTAT 1080  
1140 TTTATTAAGG TAAAAAANA AAAAAAANA AAAAAAANA AAAAAAANA 1140  
GAN 1143

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1782 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 120:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

60 CAGGCGCGCG CCGCCGACCC AGTCTGCGGT TGTGCGCGCG CCGCGCGCG CCGCCGCGCG 60  
120 CAAGGACAAA CAGACTGTCA GCGACCTCC GCGGAGTGC AATTACGTG CAGCTGCGCG 120  
180 CAGCCAGAG TTCCAGATG GTTTGCGGG GCTTCCGCTG TTCCAGAGAC TGCCTGTGG 180  
240 CCGTCACCT GCTTTACACC TTGTTAGTC TGTGCTAAT TGAATTTCT GCGTGGGCA 240  
300 TTGGCTTGG GCTGATTTCC AGTCTCCGAG TGTGCGCGGT GGTCAITGCA GTGGGCACT 300  
360 TCTTGTGCT GATTGCTTGA GTGGGTCTGA TTGGAGCTGT AATACATCAT CAGGTGTGCG 360  
420 TATTTTTHA TATGATTAAT CTGTACTTG TATTATTTGT TCAATTTTCT GTATCTTGG 420  
480 CTTGTTAGC CCGAACCAG GAGCAGAGG GTACGCTCT GAGGTGCTGT TCGAACAATA 480  
540 GCGCAAGTGC TCGAATGAC ATCCAGAAA ATCTAACTG CTGTGGGTTT CAGAGTGTGA 540  
600 ACCCAATGA CACTGTCTG GCTAGCTGTG TTAAGAGTGA CCACTGTGTC TCGCATGTG 600  
660 CTCATCAT AGGAGATAT GCTGAGAGG TTTTGAATTT TTTTGTGTC ATTTGCTGT 660  
720 TCTTGTGTTT TACAGATC CTGGGTGTTT GCGTACCTA CAGATACAG AACCAAGAG 720  
780 ACCCGGCGC TATCTTAGT GCAATCTTTT GATCAGAAA CAGGAGAT TTCTTTGCT 780  
840 ATTATGATCT TGTTCATTTT CTGTAATTTT CTGTAGCT CCAITTCOA GTTTAAGAA 840



GGAAACACTA TCGGAAAAGG TACCTTATCG ATATGCGAAT TAAATATTTT TACTATATG 900  
TTCTCTACAT GTTTTTTTCT TTCCCTGCTT GAAAAATATTT TGAACCTTCT GCTCTCTGAA 960  
GCTCTGCTGCG ACCTGCGAAT TTACTGTATT CATTCTCGCG CACTCTCCAC TCGTGGCTTT 1020  
CTTACGATTT TTAAGTCGAG AAAAAGCTTGG TATGTATACA CTGTCTGCTT TAAATGTGGA 1080  
ATCTGACGCT ACATCTGACT GGTATATATTA TATGTACGAC TGTCTCTGCT AGATATGCTC 1140  
TACTGAAAAA AGATGCGGAA TTTATTTAAA TCGAAAAATA TGAATCTCTG TTATGTGTAAG 1200  
GAAAAATCCA ATTCCCAATT TTTTCTGCTC TTTTCTGCGA AGATCTGTCT TGTATAAAAA 1260  
TGTATAGTAA AAAATGATTA TTAACTGTGA GCTCTTTATG ATTAACGAAA TGTATCTTCTAG 1320  
AAATGATTTT GTCTTACGAA ATTGTGCTTT AATTTTGAAC TTTTACAGCT AAGTGCAGAG 1380  
GAGATGCTT TTCAATGAAT GTCTTAAATG ATTAATACAT TTAACTCTAG CCTCCATCGAG 1440  
AATGACAGCA GTTTTGAATA ATCAAGAACT ATATCTAATG GATCTTGATA TTGTATTATA 1500  
ATAATTTGAA GTCTTAAAAA CTGCATTTTT AAAAGAGTAA GTATTAATCT GTTGGCCCCG 1560  
GTACAAAAA GAAATTTGAT TATCTTAAAA ATGTATTAAT ACCTCTTTCA TGAATATCTT 1620  
CAATATTTGA ACACGAACTT GTTAAACCTA AGCAATTTTG AATATGATCT CCGATATTTT 1680  
GAATATGAAA TCGATTTGTG TCGCTCTGTA TATCTGTATA AAAATTTTAA GAGACGAAAC 1740  
CTTCTCTTCT GTATCAATGT TTGAATTTAA AGAAATTAAT GG 1782

35 (2) INFORMATION FOR SEQ ID NO: 121:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 610 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (41) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

GTTCGCTGCA GATTGTGTGT GCTTTCTAGG CCGTCTGTCC TCGCCGAAAG TCGTTCAAGG 60  
TATATTTAAA AACATATGGA TCCCATAGAA GCGCTACTAC ACCAAAGTTT ACCAGAGAAT 120  
TTTGATATGGA ATGGGAGCTGA TGGGCTTCAT CATTATATTA ATCCGAGCTG CTGATTAAGG 180  
AAATTAAGCT TTGAAAGCTT CAGCGCTCTG TCGTGTGTCAT CACACACGAA TTTTACTGTGA 240  
GTATATGTA AAGAAAAGCT CAGTCTGCTT GTTAATTTGA GCAAGCGTGT TTGATATGAGG 300  
AGCGTGAAC GTCACTATAC ACTGTATATA GTACGCTTTA CTTCAATGCA TGAATTAATG 360  
GATCTGTAG ATGACATGCT ACCTGTGACT GCTTTCAGTG TGTTCCTCT CAGCGCTCTCG 420  
GCGTCTGAG CATTACTGTA GTATATTAAT TTCTATGCGAG CCAATCGAAT CAGATTTCA 480

CTGAGCGAAC GATCATTTGG AAAATATTAAC CTCAAGTTGA CAGAGCTTGG TGTATCGAAT 540  
CGAAGCACTC ACTGTATATC TACTGTCTGA TAAAGTTTAA TTAACTTTGA AAAAAAATA 600  
AAAAAAAA 610

10 (2) INFORMATION FOR SEQ ID NO: 122:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 526 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (41) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GGTACGCTTG CAGATACCGG TCCGAAATTC CCGGTGCCCC AACGTCAGG CCAAGCGTCC 60  
ACCCAGCGCT CCGGCGACCG GTTCGAGCGG AGCCGAGATG GTCAAGATGA TCAAGAGACT 120  
GAGCTCGCC ACTTTCGACA ACATCTGTGG CGTGTGCTCT TTCTTCTCTG TCGTCTCTTA 180  
TCACTAGTGG GCGCTGACA ATCCGAGAA CCGAGAAATG AAGTGCAGCT TTCTTCCGCCC 240  
CAGGTTTGA GAGCATATTC TGAAGCAGAA TGAAGGATAT GAGGCGGCTT CACACTTCAAC 300  
TTGATCCCTT CTACCAATCA CAACTATACA AGCAATGACA CCGGATTTT TCCAAACGAC 360  
TTTATTTTCC TGAAGTCTT CTTTAACTCT ATGGAACGAG AAGCTCCGAC TGAATATGCGC 420  
CGAGTATAGG GCTCTGCTTT TCTACTCCCT CCCCCAATA TAAATATATA GACTTTTATA 480  
AAAAAAAA AAAAATTTTG TCGCGGCGCC GTTACCAATC CCGCTTA 526

40 (2) INFORMATION FOR SEQ ID NO: 123:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2081 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (41) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

TGTACGCTTC CGAATATTC CCGGTTCGAC CAGCTGTCTG GCGGACATAG CGCGCTTCGG 60  
AGCCGCGCGT CCTTGTGCTC CCGAAGAGCG GCGCGCGCGG CGCGCGCGCG CCGTCCAGCA 120  
CAGTCCCGCT GCTCTCTGCT TTCTCACTCT TCGCGCGACG CTGTCTGCTT CCAAGCGCGG 180  
CGGCTTACGA GCTGCTATTC CAGAAATTC TCAAGCTTGA CCGCAACGAG ATGCAATATC 240  
ACCGAAATTT CCGTGTGCGAG CTGTGTGCGG AGAGATGCGG CAGATATGCT GACTTCCCA 300

5 AGGGCTTCC GSTRAGCGAG CGCTGCAAGG TCGCCTCTGT GCGCTTCCAG ATCCAGCTCA 360  
CTACCTCCGG AATCTTACCA CTTTCAGCA CTGTGTTTTT CTGCTGTAT ATCCAGGAAA 420  
GCTTCAGACC AGCCATCAG TATTTTGGG ATATTATTAG COTGGACAG AGATTGTTC 480  
AAGGGCCCG GATTTTAGA ATTCCTGTA TTGTACAGA AONTACCT AAAGTCTTG 540  
GGAGAGCGT TCAGAAATTT GATTACAG GTGTAAACT GTACTTCCA AGACCAAGT 600  
TTTCATGCT ATTACAGAA GTAGAGCGG CATTACAGA GATTCGCGA GTACAGAGTG 660  
TTGTATTATT TCGATAGAA ACTATGTGT GCATCCACA AACTGCCCTG GAGCTAGTTG 720  
GCCAGGAGT CGAGGTTTAC ATTGTGCTG ATGCGACTC ATCAGAGAGC ATGATGACA 780  
GGATGTTTCC CCTCGAGCGT CTGCTCBAF CCGGATCAT AGTGACCAG AGTGAGCTG 840  
20 TTCTGCTTCA GCTGTAGCT GATPAGACC ATCCAAATTT CAGGAAATTT CAGATCTTAA 900  
TTAAGGCGAG TCGTCCAGAG TCGGCTCTCC TTTCGAAAGT ATAGGACATT TCGAGAACTG 960  
GTATGCTACT CACTGTGAA GCACAGTCAG GTGAGGACT GTAGGCCAC ACAAAGCTCTT 1020  
25 CTATCTCTA CTAGATTAA ATGTATAGT CAAAAGGCG TCTTTTTTG GCGCTCTTAG 1080  
TCAAACTTAA CAGCTABAC CATTGTAGTA CCAGCTTTA GTTACAAAGG TCGAAGGCTT 1140  
30 CCGGTGTCG TTACCTTCTT TTTTGTGTA TGTGCTTTTA TTTATTAAAA AAAATTACAA 1200  
TCAGATGCC TGTTTTCTCT CTACTGTGTA CTCTGATGCT ATCTTTCCAA AGTCGAGCT 1260  
CTTGAGAGT TTCTTAATTT TGTTCACITTT AAGAAATG AGTACCAAC AATGNTTGG 1320  
CTTTTATTT ACTGTAGAT GTTATATGT TATGTGGAT GTATGCTTT TACTTTACAG 1380  
ATTGATGGA ATAGATTAT TCGATATGAA TTTACCCACA GACTCTGAA TCAATTTACC 1440  
40 CACTGCCCTC ACATGTTTGT CCACTTAGTG APTTGCAITG ATCTATCGGT ACCAATGAT 1500  
GTTGATPAT TACATPATTT TCTTGACTAT ACTGATTTCT TATTTTGGTC ACTATTACTA 1560  
45 AATCTCTGTT AATATTTCTT CTTTAACTG AAAGGCGATG GATPAGAGG GTTTGCAATG 1620  
CCATATTATT GGTGAGGCG TGTTTTACCA TCTTTGAGT ATGCTTGTCT GATATCTTT 1680  
ACCAACATCT TGATPATTA TTCTAGTCTC CACAGNTTT ACCAAGAA TAAAGCTTGG 1740  
50 GTGGATPATC ATTTTAAAT GTTCATGTTT TGTTCATPAT TTTCTTACC TACTCTCAA 1800  
ATATTTTAAI GCAAAAATG TCAGTATGA TTCTGTAGTA TTAATTTTGT GGTCAATGTT 1860  
TCTCTCGAT AATTTATTT TCAATTAATA CTHTTTAGAG GGTTTTGAAA TGTTTTCAA 1920  
55 ATATGTCGAA TGTGAACTG CTGTCTTTTA TATTAAAGTA ATTAAAGAA ATGTATTTG 1980  
ATTGAATTA TTTTGGCTC CACAGATGCG CTCTATGAGT ATTTCTCCAG GCAATCTPAT 2040  
60 ATTATTTAA GGTATPAAA TCTTGACATTT TATATCTTT C 2081

## 5 (2) INFORMATION FOR SEQ ID NO: 124:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1717 base pairs.

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

15 CCGCGCGGA CCTGACCCG CGGTGGGCTA GGGGAGGCG CGGAGCCCG GCGCGGAGC 60  
TGTGGATCTT TCATGATGAG AGATTTGGGG ACACTTCTCT CTCTGTGTG TAGTTGATAG 120  
TTTGTGTG AGAGATGCG TCACATGTC AAAACTTTTC TCGAGACCT TCCAGAGGA 180  
ATCAAGACT CCATCTGGG TATTGTACC ATCTCAAGC TAGATGCTCG AATCCAGCA 240  
AAGAGAGG AGCAAGCTCG AAGAGGCGA AGTAGTGTCT TGGCAGAGAG AAGAGCCGAG 300  
25 AGTATAGAGC GGAAGCAAGA GAGTAGGCA CATTATGTTA GTAGAAATTT CAGTGTGTT 360  
GCTTGGATG GTGAGTGTT CTGTTCACT CTCTCTTGT TTTATCGACT ATTTATCTCT 420  
GTCTTCACT GGTATACAG CCGAATATC GGTGACCCAT CACTACATGG AGATGTTTGG 480  
TCTGTGGCTG AATTTCTCT CAGTCAATTT TCGAGTCTC TTTGGTGTCT CCGCTTGTTT 540  
GTCTTAGCA AGGTGTGAA TCCATTTGG TTTCAAGATA TAGCTGACT GGCATTTTGA 600  
35 GTATCAGGA GAGAGCTCA CCAATTTGCT AGTGTACGA AATATTTCC TGACATGCTC 660  
TTCAAGCTTT TCTGAGGCG TCTTTTCTC ATTCAGGAA TGTTTGTGAG TCTCTTTTCC 720  
ATCATCTTC TCGTCACT GGTATGCTC CTGCATATGT CCTTCTCTA CTCACGTATC 780  
40 TCGTTTGAT ATCTTTGCTT CAATPAGGA ATTCGAAATCC ACCAGCGGTT GTCTAACATA 840  
GAAGGAATTT GCGCTTACTA CTTTGGGTTT GGTTTGCGCT TGGCTTTTCT CACAGCAATG 900  
45 CAGTCTCAT ATATPACAG TCGCTGCTTT TTCTCTATCC TCTTTTCTTT ATTCATATTC 960  
AGCGCAATG AAGCAAGAC CCGTGGCAA GCTATCTCT TCGAGTTTGG CCTCTCTTCC 1020  
50 TTGTGTGCT TCTTAAAGCA CAGACTCTTC CACAGACAG TCTACTTCCA GTCGGCCCTG 1080  
AGCAAGCTTA CTCTGCGAA GAGTTTCTCT TCAAGGATC CGTGCGCTCC CAATCTGAG 1140  
GCTACTGCG GTCACTGAGT TCGCTGCGAT CCAAGGCGA TGGCGCGGAT TCGAAGAGC 1200  
55 TGTGGAGCT CTTTTCCTG TTAGCTTCC GCTGCGAAG GAGGAGGA CCGCTCTCC 1260  
CAGGCGGCT CTGCTATTC CTTTCTCTCT GAGGAATGCA AATTTTCTC TCTGTGTCAC 1320  
GTAGGCGGA ATGTTTCTCG ACACAGATGT GTGATTTT ACATCACCG TGAGTCTGAA 1380

AGGACACAG GTTTCTCTGC AGCTATTTC TAGCAATTC GAGTCCCTCT GCTCGACTG 1440  
ATTGAGACAC TTGTGTTTTC TCCCTGTGCC ATTACCCCTT GCACCTTCC ATCTGACCTT 1500  
CTACACCCCT TCGATGATG GATTGTGTA TTCTACCTCT TGTATTTC GAAATTGTGA 1560  
ATTGTTCCT TTCTCTGTA AACATATCA TTGATATGAG GAGTAAAG AGTGTCCAG 1620  
TTGCTCTGAG TCACTCCCTT TATACCAAT ACTGTCTCT TTCTGTAC TCAAGTTAGG 1680  
TTTTCCTCTC TCTTGTCCA CTCGAAAA AAAAAA 1717

15 (2) INFORMATION FOR SEQ ID NO: 125:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 804 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

CGAGGCTTCC GGTGACATAG TAGTGAGAG GAGAGACACC TCCCGCAAA TCTGAGAGCT 60  
TCTTAGCTTC GACTAGGACA GTAGCCCCAG GACTCTAGT GCGCGCTTC AGTCACTGC 120  
CGGCTAGAG GAGTGCCTCT CCGCTGTTT GCTCTCTTGA TCGCGGAGAG GCTGTGTGAA 180  
ACAGCTGAC AGCAATGTGC AGAGGATGAA TTTGTTTTC ACTTACTGA TTATGAAAT 240  
ATCAACAGG TTGTGTTTTT TATCTGGGA ACATGCCAT TTCTGAGAG GATGAGAGAA 300  
TCTGTCACT TTTCCTATCC TGATTCAAT GGAATGCGAG TAGGCAACT CCTAGAAAT 360  
GTCAAGATG GAGAGCGAG TGCATCTTC AAAATTTCAG GTCTTAATC TCGAGAGGAA 420  
AGCCAGATC CTTTTGAGGC CATGAAATTT GTCCGACTC GATCTGTGC TCAATTTGAA 480  
ATTGAGTGG AATTATATGA CAGTATGCTT CAGCAGACTC CTGTAGGTGA TGTGCTGTGA 540  
TCTGATGTT AGTCAATCAC TCAATTCACA CAAAGATGT TCGACAAATT CTACAAATTT 600  
GCTTCACTAT TTGCTGTCTC TCGAGCCGAG ATGACACGAA GCGCACTGTA AATGTCAAT 660  
CGGCAATAG TGGTCTGAAA ATGTATGAAA AACTTGAAA GAGCAGTAGC AACAGACCT 720  
AATTTTGGN AATCATTAAT TGATTAATAT AATTTTAAT GGAATTGTGA AAAAAAAA 780  
AAAAAAAA AAAAAAAA AAAA 804

55 (2) INFORMATION FOR SEQ ID NO: 126:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 431 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

GGCAGACCC AGGCTCTTGA AGCCAGCTG CCGTAGAGAG GGGCTGCTGT GCGAGCTTGG 60  
GAGGGCTG GAGTGGGGCT GCGCTGATG GCGCTGATGT GGAATCACTT GCGAGACTCT 120  
GCTGGGGTGA ACTTAATTTT AGCCCTTCCC TTGTGCTCT TATGAGAGAA CAGAGAGAG 180  
GTGGGCAAGT CAGTATATTC ACAGTGGAG TGATTCCGAG CAGAGGGCTT TCTGGAGAGA 240  
GGCAGTAGAG GCAATTTCTT CAGGAGATG GTCCATTAAT TCAAGCAGAA GGCATTGCTAT 300  
TAGGTTAAT CCGAGACTTT TGTGCGCGAG CTCTGTGTTA TTAAAGGCCC TTGAGCAAGAA 360  
CTTCAAGAG GGGCGAAGAG GACTTTAG TTTTATGATT TTAACAGAGG AACCCAGAAA 420  
GGATTATTT G 431

25 (2) INFORMATION FOR SEQ ID NO: 127:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3752 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

AGGACAGAG AAGATCACTT GACTCAGAA CTAGAGATAT CGAATGACC AGACAAAAT 60  
AACTTCAGC TTCTTAAGCA TAAAGATTT CAGAGACTC TTGTGTGAAA GAGACCTGTG 120  
TAGATATGCA CATTAGAAC TGGCAGAGCA CTGAAGAAA AAGCTTTCCT TCCGAAAGAT 180  
ASTCAGAAC TTGACAAATTT CTATGAGAAA GTTCAGAGCA AATGAGATAC TGTTTTGTGC 240  
AAGTCTGTG AAGGTCAGCA CAGATTGAG GAGGCTTGC TCTTTTGGGG TCAATTCATG 300  
GATGCTTTC AGGCATTTGT TGACTGTGTA TAGAGTGG AGCCAGACTT GCTTAGAGAC 360  
CAGCCCTTGC ACCGGGAGAC TTGACTCTGT GATGAACTC ATGATGTGAC ACAGAGTTTT 420  
CGAGAGGAAA CTGAGAGAGG CGAACAGAAA CCGTTCAGCT CTGAGAGGAG TCAAGCCGAG 480  
AGCTAATTTA GATATGATGTA GATGACAGCA CTGTGTGAAA AAGACAGCTC CAGAGACTGA 540  
GCACTGCTG GAGCAGCTCT TGTAACTCT CTGTTCGAA ACAGAGCTGG CTTCAGAGAG 600  
CTTAAAGCA AGCGAGATG TTTCAGAGCA CAGTCCAGAT GCTGTGTGAG TGGCTTTCTG 660  
AACAGAGCA AAGCTTGC TTTCGGGAG GACTCTGAG ATGACAGAGA GCGCTTCAG 720  
TCTGTATATG AACCCAGTAA GAAATTCATG AAGAAATGAG AAGAAAGCT AGTGAAGCTT 780

840 AACTCAGGAG TAGCCATGGG AGAAGTATC CTGGCTGTCT GCGACCCGCA TTGCATGCA  
900 ACCATCAAC ACTGATGATC CATCTGCA GCTGCTCTGG AGGAGGTCTT GACATGGGCT  
960 AAGCAGACC AGCAGCTCTT TCGAACGGCC TTGTGAGAAC TGCTGGCTTAA TGCTGAGCTC  
1020 CTGGAGAAC TTCTGGCATG GATTCAGTGG GCTGAGACCA GCTTCATTTCA GCGGATCAG  
1080 GAGCCATCC CCGAGAGAT TGACCGAGTT AAGGCCCTTA TGCTGAGCA TCGAGCATTT  
1140 ATGGAGGAGA TGACTGCGAA AGACCTGAC GTGGACCGGG TCACGAGAC ATACAAAGG  
1200 AAAAAATAG AGCCTACTCA GCGCCCTTTC ATAGAGAAAT CCGCAGCGG AGGCGAGAA  
1260 TGCTTAATGC AGCCAAACCC TCTTCCCATG CCAATCCCTT CACAGTCTCA AGCAGAAAC  
1320 CCAAGATCA ACCAGCTTTC TGCCCGCTGG CAGCAGGTCT GCTGTGTAC ACTGAGCGG  
1380 CAAGAGAAC TGAAATGATC CTTCGATCGG CTGGAGAGAT TGAAAGAAAT TGCCAACTTT  
1440 GACTTTGATG TCTGAGGAA AAGATATAG CTTTGATGCA ATCAGAAAA GTCTCGAGTG  
1500 ATGCAATTTT TGCGGGCAT TGATAGGAC CAGGATGCGA AGATAGACCG TCAGAGATTT  
1560 ATCGATGCGA TTTTAGCATC CAGTTTCCC ACCACAGT TAGAGATGAC TGCTGTGGCT  
1620 GACATTTTGG ACCGAGATGG GATGTGTAC ATTGATTAAT ATGAAATTTT GCTGCTCTTT  
1680 CATCCCAACA AGCATGCTTA TGAGCCACA ACCGATGCGA ATAAATGCA AGATGAGTT  
1740 ACAGAGCAG TGCTCATGTC CAATATGCA AAGAGTTTC AGGTGAGCA GATGAGAG  
1800 AATAAATACC GGTTCCTTCTT CCGGAAATGAG TTGCGGAAAT CTCAGAGTTT GCGCTGTGTC  
1860 GGTATTCTGC GCAACCGTGA TGTTGCGGCT TGCTGAGAGA TGATGCGCTT TGATGAAAT  
1920 TTTATGTAAG AATGATCCCT CCGCAGCAG AGGTAGACT AACATTGAC TTAGAGAGAA  
1980 ATTCAATCTA CCGAGGGAG CATCCGAGG ATGACCCC TTCCGCTCAC GGGGTGAG  
2040 GTCCAAACCA TCTTCCCGGG CAGCTTCCC TACTGTTTCC AGCTCCAGTG CTAGTCAGAG  
2100 TAACCAAGC TGATACATCA TGCCATCTTC TCAGCCACC CCAAGCAGTG GAACCAAGT  
2160 TATCCCATCA TCAGATGAGA AGTTGAAAG AGCAACCA ACTTTTCAAT CTAGTCGAC  
2220 ATCCCTTGTCT GGTGATACA GCAATGATTT CTTCGCCGGC CTCACAGGT GCCAAACTA  
2280 ATCCGGCAGA CCTTAAGAG TCTCCCATG GCGCTGGGAG TGCGGCTGG AGTCGAGCG  
2340 GAGTGTGAGC CAGCAGCCGG CAGGAAGTG AGCTTCTGCA CTTTGACCTC TTAGAGAGCC  
2400 ATTGCTTTGT CCGACACTTC AGAAGCAGC GCTGAGGGGG GCCAGGCA CTCGAGAGA  
2460 GGGCTTAACA AAGCTTCCA AATCCCAAC ATGTCTTACA AGAACACAC TGCCCTCCCG  
2520 AGGACTCCAG GTCCGAGGG ATTAACACT CTAGCAGCC CCAAGCACT ATCCACTTTG  
2580 AATCCTGCTC CATACATG GGTATTAAT ATTGTGAGG GAGAGATTA TATTGTTAAA

2640 AGGTAAAG AATATGTTG TTATGAGCT GCTTATTTT TTTTCTTTT GTAGTTACT  
2700 ATTTTCATGT GAAATTTTAT GTAGTAAAA TTGCTCTCT GGTAAACCTG TAATGATGG  
2760 GGGCCAGAA TGAATATTT GAGAAACCA ACTGAAAGG TCAGATGACA AATGTGTAT  
2820 AAAAAAAA AGCCTATTA ATAGGTTTC TGCGGGTGC AGGTTGTAA ACCTGCTTAA  
2880 TCTTTTGA TAATCTCTAA ATGCACTTC TTATTAAC TGACTTCTA TCTGAGCAG  
2940 ATAAATATA TTAATAAAT AAGATCTCG CATGTGTTAA GGAATCTTT TTTGTAAAT  
3000 CAGCGACCC TCATTAAGCA AGAATGAGG GAGGGCTTT TTCCATTTT TAATGTTTG  
3060 TGATTTTTAG CTAAAGAGG GGAACCTCAT CTAAATACA TTTCACATG ATACAGCAA  
3120 AGAGTTTCA TGCATACTG TCTTTGATA TTGTTTCACT ACTGGGTGT TAAAGACAA  
3180 ATAGCTCTTA GAATTCAGG GTAAATGTA GTTTCAGAA AACGTCAGAA CATTTGGGT  
3240 TTTAAACTGA TTTGTTGCTC CCTATCCAG CTAGACACA GTAACTCTTG TGTTCACAG  
3300 GAGCCAGACC CTGCGCAGG GATAGGCTCG TTGTGACAT TGTGAATTC AATTTGTTT  
3360 TATCCACTTT TTTTGTAT TATTAAATG GTGATCAG TCCTCAGAAA CTGAGGAATG  
3420 AATTCACGA GCGCTGCTG AAAATGTGA GTTAAGACA ACAGTGTCT GTCTTTTAAT  
3480 GGAATTGACC AGCACTATG TTACCAATG CTGTTGCTT TGCTCTTTT TTGTCGTTA  
3540 TAAGTTCAAC TGACCAAGTG ACCATGAAA GGGGCTGCT GGGGCTCTG TTTTGTAGCT  
3600 GCTGTTCTTC AGCTCGGACC ATGTGCTGT GTGATTAAT CAATTTGTT TAATGAGGC  
3660 AGAACTGA GCTCTACCA TGACTGTTT AGAACAAGA CACACTTTG TATTAATAAT  
3720 GCTTGCATGA ACATAAANA AAAAAAANA AAAAAAANA AACTCGAGG GGGGCGGCT  
3752 ACCCAATGCG CGTATATGA TGTTAAACAA TC

(2) INFORMATION FOR SEQ ID NO: 128:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1144 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

55 TGACCTCTG CCTGCGGGC TCAGTGTG AGCTTTCTG TTTTGTGCA GTGCTGCTC 60  
GTTAAACCA GGGGCTCTG GTCCACCACT CCAATTCAGA GCTCCATTTG GTCCAGAAC 120  
CTTAGCAGG CTTTCCCTTC ACAGTCCAG CAACAGCTT GCGAGCATC GGCCTCATG 180

	GCACAGAAA CTCCTCTGCT CCTCAGCTC CCTCAGCTC GATTCAGCT GACGACTTGG	240
	GAAGAGCTTA CAAGCCCTGG GCGATTGGA GCGCCAGCAT TCGAAGAGA AGCTCGAGCC	300
5	CTTCCTCTTA TTCCGACTTT CCTCAGAGA ATTAAATTTA GCAAAAACA ACAGAACTTA	360
	GTGGGCCCC CTCATGATCA TGAATGCA GTTCTGAGA CACTCTTTTA AGGCTTTAC	420
10	TGCAGTCCC CTCGCCAGCC TCTCTCTCT TCGAAAAG AGCCAGAGAG GCGAGCTTCA	480
	GACGACTGCG TTCTTTGAGA TCTTCTTCC AATTGAGTA AGATTGAGTT TCGTCTGTGG	540
	CTTTAGAGA AAGTCTGGA CTCAGCACA AACTCTAATA AGACCTGTAC ATCTGAGAAC	600
15	CTTCCCGTCT ACTGCGTTT GACGAGCTGT CTTCGCCATG CTTTATTAT CTGTATGAC	660
	ACGATTTGA CATTAGACT AAGGAAATA TTGAGTTGA TTGAGAAATC CTGCGATGTG	720
20	ACGATTTGT TAAATTACA AGTTGGTTT TTAAATTTT CTGATTATA TCGGCCAGAA	780
	TCATATTTTA AAGCTATATG AGGACTTGT GCTGAAATA GAGTATTTT TTAAAGTAAAG	840
	CGTCTCTTGG TTAAAGACA GATTACAGA ATTGATGTA ACTTAAAGAC TGTAAATGAA	900
25	TGTAAAGAA TTGAGTTGAG AGCAATGCA TTTTCTGTGC TGTATTACT TGAAGTATGT	960
	AACATTGTGA TACCTGAACT TATTTTAAG ATGAACTGAA ATGCAATAG CCAAGCTTGG	1020
30	AGATACAGAA TTGAATGTGT ATTCTTAAA AATACAACTT TGTGTGTAC TTGGAATATA	1080
	ATGATGCTTT TTTCAGAAAA AAAAAAAAAA TCGAGGAGGG GCGCGTACG	1140
35	CAAT	1144
(2) INFORMATION FOR SEQ ID NO: 129:		
40	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1830 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
45	(1) SEQUENCE DESCRIPTION: SEQ ID NO: 129:	
	GAATGAGAG GAGGAGCTGG AGCTGTTCG TGAAGAGAGC GGCATCTTG GAGAGAGCC	60
50	AGCGTGTCT GCGCAATGCG ATGATGAGAG CAGAGCAAT CATGCGAGAC AGGCGCCAG	120
	GATGAGATT CTGACCAAAA ACCTGCAAGT CAGAGAGAG ATGCGCCAGG GTGCTCCAG	180
55	CGTCTACGC CTCTGCGAGC GCGCGGTGGA TGGGAGCTC TGAACAGCA AATGCCCGAC	240
	GCTGAGCGGC GCGCTGTGGA GCTGAGATT GAGAGAGAGC AGGAGAGAGC GCTGCGCTTC	300
	TCCAGGAGTG GCGCAAGCT TCCGAGAGCG CCGTTCGCG GACTGAGCA GCGTCTCTCC	360
60	TGCTCTCTTC GCGAGAGAGC CTGAGCAGCC TCGTGTGAT GAGAGAGTGG CTGATTGCGC	420

	CTGGGCGGCG CAACTCTTCA GAGCGTGC AGAGCTGGA GGGGTGCTGG AGACCAATAG	480
	AGCTGATGAG AGGAGCTGAT GCTTGAGCTT GAGCTCTGCG GTCCCGAGA CCGAGAGAAA	540
5	CTGATGAGAG GCGAGTGGG GCGAGCGGCG TCCCTGGAGA TGGCTCCGCT GCACTTTTGA	600
	AACCGCGTT TCCCTCAAG TCCGATTC AGGTGACAGC AGGTGATCCG TCCCTCTCAT	660
10	CTTACCTTCC AGGTGACACC TTAACCTGTG CTAACTGCT TGTGAGACT GAAAAAGCT	720
	TGAGCTGTGC GGGAAAGAG AGAGGGGCG TCAATGAGC CTGTACAG AGGATCCCGG	780
	AGAGGAGAG GAGCTGTGGA GATGAGCGC CGTGAAATCG GGGCAGAGC CCGTGGGAGC	840
15	GAGATTGCT CTGACTTAT TTATATGCA TGAATCTCT GATTATTTT GGGATTTTTT	900
	GTGTGTGTG TTGTCAGAT TTGTTTTTC TAAAGTTGCG TGAATTATTA TTGAGCAATT	960
20	TGACTTTCA AAGAAAGTAT GTGTCTAAC AGGAGAGCA CAGAAGTAG TATTGAGAGC	1020
	TGTTCCTGCT TCTACTAAA AAAAAAAAAA ACATTAATTA TTGAAAAATT	1080
25	AAAAAAAAAT AATTGTTCT GTTTTATAT ATTAGCTTG TAAAGTGTG TTGTAGAGTA	1140
	TGAGCTGTA TTCTTTCCCG GAGCTTCAT TCTCAGCGGG TTGGCGGTTG TTAAAGCTG	1200
	CTCTCTTGA GTAGTGCCT AGAAGGCGCT GAGAGGTGCG CAGCGAGGT TGAAGCTGGA	1260
30	GGGATGAGAG CCGGAGCTGA GGTGCGGTGT GAGAGGGTT AAGAGGTGAC TGGGAAAGC	1320
	CGGGGAGTGG CTCTGTGAT TTATTTTCTT GATGTATCT TCTCAGAGCA GGGCAATTGG	1380
35	GAGTGCACA GCGAGAGAGC AGGAGGAGC CCGTCTGCT GGGAGAGAG GAGCCAGCA	1440
	AGCCCCCTGG GAGTTTGTTC CCGGAGCTT GCGATGCGCT TCGAGGAAA GTGAGAGCTG	1500
	GGAGGAGAGC GGGGAGAGAG CTGTACCCC TGGAGGTGCG TGAATTCAAG TGGGGGCTCC	1560
40	CTGCTKCCC GAGGCTGGG AGCTTGAAGC CCGTCCGACA TCTGCAATCC GAGCTTCCCG	1620
	CGTTCAGAGG TGCGCTTCCC TCTGTTCGCG CAGGATGAC GAGGGCAGAGC AGTGGCTTGG	1680
	TGAGTGTATC TTGCATCAGA GACAAAGAG GAGCGGCTTT AGCCGTCTG CCGGAAATGG	1740
45	GGAGAGCCC AGGAGAGAGC GATTGTGAG TGTATTACT TAAATGTAC AGATTCTCT	1800
	CGTTAAATTC TTGATGAGTT TTTTATTAAT	1830
(2) INFORMATION FOR SEQ ID NO: 130:		
50	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1864 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
55		
60		

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

5 GGCCTCCCG ATGGCGACCC CAGCTCTGCG CCGACACACA CGGCTCTGCG TCGCAGACTT 60  
TGTAGTGTAT AAGCTGAGCG AGAGGGTTTA TGTCTGTGGA GCTGGCCCCG GGGAGGGCCC 120  
AGCAGCTGAC CCGCTGCACC AAGCATGCG GCGACGCGGA GATGAGTTGCG AGACCCGCTT 180  
CGCGCGACCC TTCTCTGATC TCGCGGCTCA GCTGCATGTC ACCCGAGGCT CAGCCCAACA 240  
ACGCTTCACC CAGGCTCTCG ATGACTCTTT TCGAGGGGGC CCGACTGCG GCGCCCTTGT 300  
AGCCTTCCTT GTCTTTTGGG CTGCACTGTC TCGTGAGGT GTCAACAAG AGATGGAAC 360  
ACTGCTGGGA CAGTCTCAGG ATGCGATGCT GCGCTACCTG GAGAGCGGCG TCGCTGACTG 420  
GATCAGAGCG AATGGGGGCT GGTATTCGCA GATCCTGAA GCTGAGATCG CTGATGAGT 480  
AATTTGCACT GAAATTTTAA CGGACTGTGA CTCTGCTGCA AGTTCCCGAG ATCTTGAGGA 540  
GCTCGAGCT ATCANAGCTC CAGTCAAGGA GATGAGGAA GAAGCTGAGA AGCTAAAGGA 600  
GCTACAGAC GAGGTAGAGA AGCAGATGAA TATGAGTTCA CTTGCAAGCA ATGCTGGCCC 660  
GATGATCATG TCCATTGAGG AGAGATGGA GCTGATGCC GATTCCATCT ATGTTGGCAA 720  
TGTGCACTAT GGTGCAACAG CAGAGAGCT GGAAGCTCAC TTTCATGCT GTGTTCACT 780  
CAACCTGTT ACCATCTCT GTGCAATTT TATGTGGCAT CCGAAGGCT TTGCTATAT 840  
AGATTTCTCA GACAAAGAT CAGTGAAGC TTCTTTGGCC TTAGATGAT CCGTATTTAG 900  
AGGAAGCAA ATCAGATGA TCCCAAAAG AACCAAGGA CAGGCTATCA GCACACAGGA 960  
CGCGGGTTTT CGAGAGGCC GGTACGGCG CCGAGCGACC AACTACAGCA GCTGCCCTC 1020  
TCGATTTTAC ATGTGTTTAA ACACAGGCC CCGGGTTCG GTCTACAGCG GCGGGCTAG 1080  
AGCGACATCA TCGTATTTCC CTACTAANA AAGTGTGTA TTAGAGGAG AGAGAGGAAA 1140  
AAGAGGGA AGAGGAAA AAAAAAGAT TAAAAAANA AAAAAAANA ACGAGAGTGG 1200  
ACCTTGATCG AAAAAAATA TTTTAAAAA AAAGATATA CTGTGAGG GCGGAGATC 1260  
CCATAACTAA CTGTGAGGA GCGACTGCT TTGGGGAGTA GCGGAGGCC CAGGGAGTGG 1320  
GGAGGGGGC TCGTATTTCA CTCTGGGAG TCGCATGGA CACGTCTCAA CTGCGCAACT 1380  
GCTTCCCAT GTTTTCCCTCG CCGACCCAC CCTCTTCTC GGGTCCCTG CCGCTCCAGA 1440  
TTGCTCTGTC ATCATTTTGG TTCTCTTTTG TGTATTTTAT TCTGTTTGA GTGCTTTCT 1500  
TTGAGGTTT CTGTAGCCCG AAGATCTCG TTCCGCTCCC AGCGGCTCCA GTGTAAATTC 1560  
CCCTTCCCCC TGGGGAAATG CACTACCTTG TTTTGGGGGG TTTAGGGGTG TTTTGTGTTT 1620  
TCAGTGTGTT TGTTTTTTGG TTTTATTTAT TTTCCTTTGCT TTTTATTTGG 1680  
AGCGAATCGG AGGAAGTGG AACAGGGAG TGGAGGTGG ATTTTGTGTA TTTTATTAGC 1740

TCATTCCAG CGTGGGAAT TTTTTTTTAA TATGTGTCAT GAATAAGCTT GTTTTGTAAA 1800  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1860  
5 AAAA 1864

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2041 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 131:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

20 GGCACGAGCG CCGCGAGCG CCGTGGACCC CGCGCGCTCC CGGGGATGGT GAGCAAGGCG 60  
CTGCTGGCGC TCGTGTCTGC GTCAACCGC AGGAGGATGA AGCTGCTGCT GGCATCGCC 120  
TTGCTGGGCT ACCTGCGCTC TGTTTGGGCG AACTTCTGTA ATATGAGTTC TATCCAGAAA 180  
AATGCTGAC TAAATTTTGA AAGCAGATTT GAGAGATGG TTGAACCTCT AAGAGAGAAA 240  
ATCAGAGATT TAGAAAAAG CTTTACCCAG AATATCCAC CAGTAAAGTT TTTATCAGAA 300  
AAGATTCGGA AAGAAATTTT GATTAACAGGA GCGCGAGGGT TCGTGGGCTC CCATCTAATCT 360  
GACAACTCA TGAATGAGCG CGAGAGGCTG ACCGTGTGCG ACAATTTCTT CAGCGGGCAGG 420  
AAGAGAAAG TGGAGCACTG GATGAGCAT GAGAGCTTGG AGTTGATTTA CCACGAGCTTG 480  
TGGAGCCCT CTACATCGAG GTTACACAGA TATACCATCT GGCATCTCCA GCTCTCCCTC 540  
CAACTACTAT GTATATCTCT ATCAGACAT TAAAGAGCAA TACGATGGG ACAATTAACA 600  
TGTTCGGGCT GGCMAAGCA GTGCTGCCC GTCTGCTGCT GCGCTCCACA TCGGAGGTGT 660  
ATGAGATCC TGAAGTCCAC CCTCAAACTG AGGATTTACTG GCGCGAGCTG AATCCAAATAG 720  
GACCTCGGCG CTCTACGAT GAGCGAAAC GTTGTGAGA GACCATGTGC TATGCTTACA 780  
TGNACAGGA AGCGTGGAA GTGCGAGTGG CCGAATCTT CACACCTTT GCGCCACGCA 840  
TGCACATGAA CGATGGGCA GTAGTCAGCA ACTTCATCT CCAAGCGCTC CAGGGGAGC 900  
CACTCAGGT ATACGATCC GCGTCTCAGA CAGGGGCTTT CCAATGATCT AGCGATCTAG 960  
TGAATGGCT CTGCGCTCTC ATGACAGCA AGTCAAGCAG CCGGCTCAC CTGGGGAAC 1020  
CAGAGAACA CACATCTTA GATTTGCTC AGTTAATTA AACTTGTGT GTTACCGGAA 1080  
GTGAATTTCA GTTTCTCTCC GAGCGCAGG ATGACCCACA GAAAGAGAAA CCGACATCA 1140  
AAGAGCAA CTTGATCTG GGTGGGAGC CCGTGTCCC GCTGAGGAA GTTTTAAACA 1200  
AAGCAATTA CTACTCTGT AAGAACTCG AGTACCAAGC AATATATCAG TACATCCCA 1260

1320 AACCAAGCC TCCAGAAATA AAGAAAGAC GAACTGCCA GAGTGAAGT CCTCACTTT  
1380 AAGACAGAG ACTACAGTAS TACCTGAT GGAATGATTT TTGGCTTTT TTCTGTCTC  
1440 GTTAAAGAA AAGATTAAAC AGCTGTATG AAGAAAGAC TGAATTTCA TTCTGAAGCT  
1500 TCCATTAAAG AATGAAATG GCTTAAAGC TCCCTCATA AAGATCGAA TTTCGCTCG  
1560 GACTTTTGA ATCTCTCTT TTATGTAAA TACGTAAAT GAACTCTGC GTATTTTCA  
1620 GTTTTATAT CTTCCTCTGA GAGCAATAT TGTGACTCT GTTGAGATT TTATTTACTG  
1680 GTTCTTTCT GAACTGAAA AAGAACTTA ACCCGAGCA AAAATGCCA TTTTATTTAT  
1740 AAAATGCGT ACTTAAATA TGAATCTTA TACTATCAT AAGAAAAAT CTTACAGTA  
1800 TTGTCAAGAG GTGTGAGCC GGCATTGAT TTAGCCAGA TAAAGAAAT CTGTGAGAA  
1860 GCTTATATG TCTCTTTAA TTCAAGTTT TTCCAGTCT TACTTTTGA TTCCAACTT  
1920 GACTTGAATA TATCTCTCT GTTCAATATC AAGATATTT GAAATCACTA CTGTCTTTG  
1980 CTGCGTATCT GAGCGAGAG GAGTTGAGG GGCACAAAT TAACTATCT TTGGTTAAC  
2040 AAGGTAAAT ATCTATTTT AATAAAAAT TGAATCTAC GAAAAAATA AAAAAAATA  
2041 A

(2) INFORMATION FOR SEQ ID NO: 132:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2012 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

60 TACCAAGCTG CAAGATCTA CTATATCATG GAGAGAGAG TGAATGAGAA CTATGCGCT  
120 GACCGAGCT GAGAACGGA ATGCAAGAC GAGTCTGAA AAGAACTTA TGTATCAAT  
180 TTCTGAGAA AAGAGATAG GCTCTGAGT TCCAAATCA AAAAACTCT ATTCAAGAAA  
240 TAACTGAGG GTAACTTAG GATCCCTAG GAAAGAGAG GACCAAGAA ACACTTGGA  
300 ATCTTGGTC GACTATACA AAGTAAATT GGTGATACC TGAATGAT ATTCAAAAT  
360 AATGCAAGC GCGCTTACTC TGTCAATCT CATGAGTGC TGAATCTAC TACTCTCTG  
420 CCACTGGTG CTGAGCTAG TGAATGATC ACTTATCAT GAAATGCC AAGAGATCT  
480 GCGCTGAGC GAAATGACT GCTTGTCTT CTGTAGCTA TTATCTGCA GTGATTTCA  
540 TGAAGAAAT GTATATGAG CTGTGAGCC CTTGAGTAT CTGCAAGAG GCGATCTAG

600 NAGCCCATG GAGAGAGAG TGAATGATC GGAATTTTG CATCTGTT CTGATTTT  
660 GATGAAAAA AGCTGTGTA TTGAGAGAA AATGTGAAA CCAATGAGT GAAATCTCA  
720 GCGATTTTA ACTTACAGA TGAATCTT TTGAGAGAA ATAAATGCA TGCATCAAT  
780 GGAATCTT ATGCCACT TGAAGCTCT ACGATGACC AAGAGAGAG AGTGGCTAG  
840 TAACTGTAG CAAATGAGC AATGTGAT CTACAGCA TCCACTTCA TCCAGAGAC  
900 TTCTCTATC GAAATGAGG GAACTAGCC GCAATGTAG TGAATGAT CCGAGAGCT  
960 TTGAGAGTG TGAATGAT GGCAGAGAC CTGTCACTG TTTTCTCTG AACAGAGAC  
1020 ACTGAGCATG TCAATGTAG CATGAGAGC CTCTCACTG TTTTCTCTG AACAGAGAC  
1080 TTAAAGCTC TCAAGCTAT CACCAAGAG ACTGAAAAA GATGCGCCC GAGAGAGAT  
1140 GAGAGAGCA ATGTGAATG GCTGAGCAT GAAATGCCA TAAAGAAAT TGAATCTG  
1200 GCTCTGTTT TGTTCGCAAT TATGTCTAC CTTCCTCTG TTCTCTGAG ACTTGTGAG  
1260 GTGTTTGTG ACCAAGATG AAGAGAGAG CTAGAGAGC ATAGAGATC CATCTGAT  
1320 GAGAGCTCA AGCTCTCTG TTTCAGAGC TAACTCTAG AGCTGAGAA TATCTCTAG  
1380 AAGCAATCT GTATGAGCT CCGAGAGCC CATGAGTAG TCAATAGCC GAGATGAAA  
1440 GCGCATAGG TGTGAGAAA GCAAGAGAG CAATCAAGT TATGTGATA TTCTCTCTT  
1500 TATTAATTT ACGTGAAT AATGAAAT GACTTTTCT TTAGTTTCT TGTCTAGAT  
1560 GGCAGCTAG CATTAAGGA GTACTTTT ATCTATATC GAAATTTCA ACACTTCAAT  
1620 TATATCTCT TCTGAGCTT GAAAGTATT GAAATTTCA GAAATGATC CTCTCAGAA  
1680 AATAGAGCC AAGAGAAAA CTATGATTT GGTTTTAC TTCTTTAG GACTGAGAA  
1740 ATTCACTTT GAACTGAGC CAATGAGCT GTTAAATTA CCGACATTA AACTTAAAGC  
1800 TAAAGATTA GCTGTATG GAAATGAG GTAGCTGAG TATGTGAAT CCAATTTGAA  
1860 TTGTATCT CTTCGAGT GAACTCTT GAAAGAGAG TGAATGAT GTTCTGCCA  
1920 TGAAGATTA CAATCTCTG AGCTAGAGC TCTCAGAAA ACCAGATCT CCAATCTT  
1980 AACCTGAGC ACTGAGAG AATGTGAAT ACCTCTCAT GTTTTGAACA GAAAACTTA  
2012 TCCATTAAAG TACTGTTAG AACATGAAA AA

(2) INFORMATION FOR SEQ ID NO: 133:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1669 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

55  
60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

5 GAGCAGTATT TTACCACTAT TGTATACAGT ATGTTACAGT TCATGTTAGG AAGTCAGAAA 60  
AGACTTTGTT TGTCTTTGTT CTGCGATGTT GAGTCATGTT TTGTGGGCTC TTGCATGGCA 120  
CAATTACCTG TTGCTGCTGC CAGATGTGCA GGCCGAGTCT AGCTGACAC ATCTACCCG 180  
10 AGGAGAGCC TGTTCCTCAT TTCTTCACTC TCCCTCCGCC ATATAGCAAC TCTCCAGGT 240  
TTAGATTACC GTTTTCGACG ACGATTAC CAAMAATGCC CCACACAGGT TTATTTACTG 300  
TTATATACTA TACTTTTAC AGTACAGACC CTAAATTTTA TTATTTGTTG CTCCGCCAT 360  
CTGATACCAA ATGTTTAAAG TTGTTTGAAA TCCAAACATG GTAGTGTTCA TGGGTAAATA 420  
TTTTCTAGGC TATGTAGAG TTAGAGGCC ATAGCATAGA AGTAATCAG TAGCATCTGA 480  
20 GACTGTGGA GGCACTAGGG CCTCTCTGGG CCTACAGACC TCACCTCCGC AGCCTCACT 540  
TGCCTCTCTC TCACACTGCC ATCAGGCGTG TTAGTGGCAC CTGTATGAGG CCAGTGTGC 600  
GTCCAGGGGA ACAGACAGG TTAAATGGTC TCCCTAGAAC TCATGAGTC AGTTTAAATC 660  
ATCCATGAC ATGAGTTTCA TTATGTGTTT ATATAGCTTT CTTAGACATA CCMAACATC 720  
ATTCAAAAT CAGTAAATTT ATTCAGTTT TGTGTTTAGA AGCTAAGTA TGTGTAGCTG 780  
30 GAAACAAA TGAAGTGT TTCTCTCTG TTAACTCAGA GTGTGCAATT ACHATGTGT 840  
GGATTAATTC ATGTTCCAGG GGCCCTTGGC ATCTCCCATG GACTGATTC CAGGAGAAA 900  
AGCCCAAGG GAAACCCAG ATTCCTTTCG ACTAGATGTG GGAAGAGCC CATGGAGA 960  
35 TATGAGTCC TGTGAATTC AGTGTGTGT GTGGCTCTTT GTTAGCACTC ATGTTGCAAT 1020  
GGTGTTAGGA GGCTCCCAT CTACCCCTTTA CATGATGAG GACACAGTGT CTGTGAGAT 1080  
40 TAACTTTGG ACACAGTGG TTAGCCCTGA GAAATGAGA GGCCCTGCTT GGACCCAGGG 1140  
AGAGAGCCA GTGACACAGG CAGAGCGGTG CAGCCCTCTT TCCCTTCCAT TTGAGCGAGG 1200  
45 TGTGCCAGG AGCTGCCCG CTAACTCTG CTGAAGCAT AGTGACATT GTTTTGGGG 1260  
CTTATCTCTG ATACATGCTG GAGCCCTGCC TCTCCACTGC TAGATGGAC CTGGATCTC 1320  
TCATCTACT CTTAGTCTGT CAGTTTCTAC GTGTGAGAG CAAGCTGTGT GGCCAGTGT 1380  
50 CTGTGACGT CTGTAGCACT TAAAMAATA TTCCAGGGTT CCCTGGAAA CCAGTCCAG 1440  
GGTTCCTATG ATCTGTAGTT TCTAAGTGA TTATPACTGG TTTTGGGTAC CTGAATTTTG 1500  
ATGCGTTAGC CTTAATATA GTCTGGCGTG ATCATGTAGA ATCTTTTCTG GTGACAGAT 1560  
55 CATAAAGTTC TATGAGGAG TTCTATCAAG GCATCCATGT CAGTGTGCT ATGCTGTGTA 1620  
CAACTGAGA TTTTGAAT TAAATTTTG TCNTAAMAAA AAAAAAAAA 1669

(2) INFORMATION FOR SEQ ID NO: 134:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1565 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

5 CACTTTTCT ATATACCTTA AGTCATACCC CTCCTTTAGT TACCTGCCAA ACTCTGCT 60  
15 TGGTTTATAT TCCAGTTAAC ACAGTTACAA AGCTGTATAG GTGTCTTTT TTCTTTGTA 120  
ACGGAATGTG TAAATCAAG TATATACATT GTGTGATGTT CTGTGTTCTG GAGTTTCATG 180  
20 AGAATTACA CATGCAATC AGTGTCTGT ATAGATCTCC CTACCTTTGT GAATTCATCT 240  
GTTAACCCCT CTTCCTTGA GAGAGCACCG GCGATGCTGG TTAACTCTTT GTCTTTTCTC 300  
TCTCTCTTAC TGGTATTTCT TGAATTACG ACAGACTGCT CAGCTGGTGT GCTTTATCAT 360  
25 GAATATGTG TGTGACCTTG CAGTTCTTCC ACAGTTGACG AAACAGTCC TACCTTCACT 420  
GACCAAAAT TAGGAGGA AAGACAGTT TTAAACGA TCCATCTTT AAGAGCGAA 480  
30 ACCGATGTGT CTATGTGCT GCACCTGCT GTTCTACTTC TCAATTCAGA CGTGTGTGAA 540  
CGATCATTTT TCACATACG GTGAGTCT CACGATACC CTTCCTGTTG TTTGTTAGC 600  
ATTEAAATCG AGATTTTGA TTGTGAATAT ATACACAGT GTTTTCCAC TGTATTTTAT 660  
35 TTCCAAAAGT TGAAGACTGC TTCTCTTACC TTTTGCAAA TAATGTATAT TCCATATG 720  
ATTCTCAAG ACTTGATAT GTGAACCTA TTAAACCTAG AAATGTATT CATCCTTTCA 780  
TCACTGTGGC CTGATGTCCC CAGCCCTCT CTCTCTTTT TTTAGATGAG ATTAGCACA 840  
40 CTCTCAATTA TTTAAACATG CACATTTCT TCAGTATGTA TGTGTAGGCC ATCTGAGCTC 900  
ATAGCTGAT CAGTACCCAG TTTCTATGCTG TGTCAATCAC ACTCACTACT TAATATGCTC 960  
45 ATGCTGAAA TGTGAGGGA AAGTATACC ATGTGTGTCT GGAAGCAT TAACATTGTA 1020  
CATTTTAAA TACTGTGATT CTGTACATTT TCTGATTTT GTTTGTTTTT ACAGAAAAA 1080  
AAAAAAGT CATTAAGCA TCGAAGAGC AAGAGTTTA CTATGATCC TTAGGGTCT 1140  
50 CTGACCTTGG CTGGCAATA GACCTACAG GCCAATTTAA TTTAGGAG TAAATATTTT 1200  
TCAAAAGCA ATTTTTTTC TGTATTTCT GTATGAACCT GCCATATCA TGAATAGAA 1260  
55 GCGAGACCA TAAAGAGAA AAGAGTGTAT GTTCTGTAT GTTCATGTA ACCTAAGAA 1320  
ACAGTGTGGA GGCAGGCGG ATCAGCGAA CTCTAGGGA CTGTGTGCT TTGGAAGCA 1380  
TCCATACCTG CATTTGCAI TCTTGTATG TATCTATAT GCCAAGACA AACTATTTCA 1440



TCATTATATG TAAATACAC TTTCCTCCAG ACCATACATA AAGTTTCCTG GATATATAT 1500  
CTTCAGATGG CATTAAAAAT TACTAGATGG CATCAATTGA AAAAAAAAAA 1560  
CTCGA 1565

10 (2) INFORMATION FOR SEQ ID NO: 135:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2007 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

20 TCATAAAGC CCGTTATAC CCGCTTGG CAGCAAGAT CCCCCCGAG CTCACAGCT 60  
GATTTGTGC CAGCTGAC AAATTCCTA GCGACACTT GCGTCAGTT CAGATTGCA 120  
GCTGTGTGG TGTGGGAGC AGCAGAGGC AAATCTCCAG CCAACAGCA GAGCTGAG 180  
AGGACTTGA GCTACGTGC CTGTGAGAC CCCCAGGCTT TGTCAAGCA GGTGTGAG 240  
CTTCCCGAAA CTCAGAGGT ATTGATTTG GGGAGACTG AATATTGCA CTTAACTGT 300  
GAGCATTC TGAATTTAC TTCTTTATG CTTCGCCCTC AAGTTCTCT TGTCAATAC 360  
ACACACACC TTGCTCCAG ATCAACAGAC ACCCTCAGG CTCACAGTAT GGGACAGCT 420  
GATTTGGAG TCCCTTTCTG TTGGCTTGA GAACTTCCT GCTTCTGTG GCTCCACTG 480  
CAGAGCAGCT CCGAGGTGG GACTCCGATT GCGCTCCAG CAGCTCTGG ACCGCAAG 540  
GCGGCGCTC TGAACAGTC GTTAAACA CACCTCTCTC TTGCGCTCT GCGATCTCT 600  
TGTCCAAATG GTCTCTCC TGGCAGATTG AATGGGGAAA GCTGCTGGCA CAGAAAAGAG 660  
AAGCGATCCC GCGTAGAGCT TGAAGAAATG CTGAGCCCG CTCACAGAG ATTAATCACT 720  
GGGAGGTTT TCAAGTCA ACATCATCT GCGTGTATTG GGGGCGAGT GTTCACACA 780  
AAGCTTCA ACTCAAAAC CAGTCGGGC TGTCTCTCT CTTCACAG CTCGGGGAA 840  
AAGAACTCC CTCCTCTCTC ACTGATTCG AAGTGTGTT GAAATGTGT GAGCAGTGG 900  
ACTTTTTC TCTTTCTCT GATGAGCA CAGTGCAGAT GCAATCCG CATTAACTT 960  
TCAAGTGAAT TTCTTCTCT GATCAAGAT GTTGTGCC CTTTAAAGG GAAAAGAT 1020  
AAGCTAAGG TGTCCAGGC ACTGTGTGAG GCGTTTAT ATGAATCTC GTTAAAGTGA 1080  
GACTAAGGA TGAAGATAT TCTTTAATA AGGCCCCGAA GTAAAGATA AACGAAAA 1140  
CTTAAAGTG AAGAGTGTG TGTCAAGAT CAAAGTAG AGTCCCTCA GTCTGATGT 1200  
TGTCTCAAG GACTTATCC CTACAAATTT CTCCTATCC ATACTCTCC TTTCACCCA 1260

CGATGTCTC CCGTCACTC CTCAGATGT CAGAGGGTGA ACCAGATCC TTACAGAT 1320  
TGGAGACAA TGAATATGT GATGTGTGA TTCTCTTAA AAACTTAG GAGCTTTGC 1380  
TACCTCTCC TGTGAGTT GTTTGGCAT TCAATTAATA AGCGACATC TCATATTTA 1440  
TACAGGTT GCGCTGTGTG TGTCCAGG TGTATACAA TTTCAGAGG TCCCTGTCTC 1500  
CTTACGCTT TTAAGACAA ACCGATCAT CCGCATGTA ACTGAGATC TTCTTATGCT 1560  
TCTAGTGTG TGGCAGTGA TCACAGAG GATTATTT ATGCATCT TACAGCATG 1620  
TTCAAGAGG GCTGATGTA ATTGTGAG GATTATGGA AAGGAGAGG GAGAGGAAT 1680  
TCATATTAT TTATTTAT ATTATTAAT TTACATCTT CTTATGTGG TATCAAGCT 1740  
GATTAAGAC TGAAGCAT AAATTAATCC GTTCTCTCT CTCCTTGC TTCTTTT 1800  
TTTTTTTTT AATTAGCAT AACATTTT TGTTCATA GATATTTG ATTGTGTG 1860  
TAAATCTG ATTAAGGTT CTGTTTTAA AGGTGATTT TCATCTCTT GGGACAGTG 1920  
GTCCAGAA CATCATCT GTAAAGAC ACAGTAGAG ATCTCTCT GATTTCAAA 1980  
AATTTTTC TGTATGAT TAAAGT 2007

30 (2) INFORMATION FOR SEQ ID NO: 136:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1291 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

40 CTTTAAAGC TCCCTCTCA CACATGCA TATCAGTGG TTTCATGTT AAAAACCAA 60  
GTAGCTGAG TTCTACTTA ATGTCACTG AGATTGCAT TGAATCAGC CATTAATTT 120  
TTTCTATTT TTAAGTGTG GCGCTTACT TTTTAAATG AATTAAGAA CTCACAGT 180  
GTTTATTTT CTACATGAC TTAGGTTTA GGAATCTTA CCAATGTTA TCATTAATC 240  
AATCTAATG GTTACTGAA ACAAATATG TAACTTTCA TTAGTGATT ATTAAAGTT 300  
ATAGAGTGG TTTCAGAAA ACATCTCTC ACATGTGAC TTCCATACA AATCATGTA 360  
TCAATAGTT ATGCCATGA AAGCAGAT GTTGTTTCA AAATTAATCT AGTTTCTGT 420  
ACATTAAT TTGAGAGGT GACACATGC TCTTTCCAG TCTTCTTCA TGTCAAGTTT 480  
CTGATAGCC ACTATTGCA AACAGTCT GTCAATCAC AAATGTGTA AATTTCTGT 540  
AATTAATTT GTCTATTTG TAAATATGA ACTAAACTT TTGGCAGATC AGCAACATTT 600

660 GCTGACCTCG TTTTTHAAGC TAACTGTAT TCTTACTAAT GTTCTATCA ACAATGGATT  
720 TCTAATATAT GCTGTCTATT TCTAATGTTC ACATTCATAT TTGAGGTTC TATCTTATTT  
780 TAAAGAGAA CAGACTTCTC AAAAAATCTT CAGAACGAGC TTATTATTGA AATATCGAAA  
840 TATTGAATA ACCCGGTGG GTTAGATATAC TCACTGTGTC ACCAAGTGGG ACATTTGCAAT  
900 GGACTGGGG CTTAAGGAGC TTGAAGAGCA CTGTATAGTA TATCTGTAAA ATGAGCCOAT  
960 CCCCACTTGA ATGTATAGT GATTAAGACC ACCTTTACCA CCCAATATAC ACCACCGGAG  
1020 GCCCATTAAC CAACTTGCTT CTGGTTCAT TTTCTTTTCT TCAATTTGTA TGTCTAGATT  
1080 CAAATGTGT GTTCTACACT GTTACAGGCT TCTCTTTTGT TTGATTAAG ATTTAGTCC  
1140 TACTTTTGTG TGGACACATT AGAATATATCA GAGACCAAAA TAGAAGATT TGTGTGTAGA  
1200 TATTTTTCAG AAGTCAGGAC ATTGTGGCA AATCATTTAT TTGCTTTT TAAAAATTCAT  
1260 TTAAGCAGTT CAGAGAGTAG ACTACTCAGA AATATTTTTC AATATTTGT CTAGAGGTTC  
1291 AATATTTTTT AATCATATT GAATCAATA A

(2) INFORMATION FOR SEQ ID NO: 137:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1906 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

60 GGCACGAGCA CCTACTTTTG TACAGACCA TGGTGTGTC CAACTTAAA CCACAGTGAT  
120 AATTTTGGAT GCTTTGTCTG CAATCTTGAC TTGTTTTC ACATATCAAT TACAGACTTC  
180 AATTTGTGAA TCTTTTAAAC ATCTTGATA AGATCAATCA GTTTTGAGG GTTACTGATT  
240 GTAAATGAAT TCAGCTAGT TGTCTGATA AGATCAATCA GTTTTGAGG GTTACTGATT  
300 TTCTCTTCC CTCTTAGTTT TTTACCAAT ATATGGAGAA GAGTAATGCT CAATCTAAC  
360 AATTTGTTTT ANTGTGTTAA TAAAGTGTCT GGGCAGTGTG GCAGCAATCC TACTTAGTGT  
420 CATTAAGACA AATACTTAC ATAGCTTTCT TAAATATAG GAATGACATT ACATTTTAG  
480 GAGAAAGTAA GTTGTCTTTC ACCGCTTACT TAAATCTTTT CCATATATAG TGAATCAAAC  
540 TTTTGAAAT GGAATCTTAC TATTTGATA GAATGTGTA TGTAAATAT ACATATATAC  
600 ATAAGCATAT ATGTGTGTGT GTGTGTGTAT ATATATATAT ATGCTGCTG TGAATCTGA  
660 CTACACAACA TAAATCACTT TTTAAATTC ACCAAGCGGT AGTGTGACAC GTGTGATTATC  
720 GTTTTGAGGC TGAATCGTT ATTAATGTTT TATTTAGGTT TTAATCCAG TAGCAAGGCA

780 TTTTAAGTAA GTTGCACCTTA CATGATTATT GTGATTTTAA ACTAAGAATA AAGCTGCAT  
840 TTTCAAGAT AATTCGAAT TCTGTGTGTT GAATTAACAA CCMAATACT GAATCTGATG  
900 TACATACAGG TTTTACAGG AAGAGATGCT ATATTTTACA ATTTCGAGAT TTAATAACCA  
960 GGGTACCCA GAAAAAGTGA CTTCATTAACA TGGTACCAAT AAGTAAGGCA TGTCTCTCTG  
1020 GTTTCTTTTT CCACATTTCA AGATTTTAACT TTCTCAGGTT ATTATATCAA ATTTTGTAT  
1080 AAGTACCA ATAGATTTTT TAGGTTTAAA CAAAGATGG GGGTTTGTG GAGTGTTTAA  
1140 TGTATGGG ATTTTATGTA GATAGAGCC TTGTTCTGTC ATTGAAATGT TTCGTATATT  
1200 TTTGTTTAC AGTTAATCTT CCTCCGCCAA GTTTGCTATT CAATCAACT GCCTGAATGA  
1260 CAATCTAGT AGTCTGATGT ATTTTCTGA GGAATAGTTT GTGATTCGA TGCAGGTGTC  
1320 TTCAATACA TTACCTCTAC ACTGCAGAG AAGCAAACT CTTTATTAG AATTAATCA  
1380 CATGTGATG GCGAAATAG TTCTGAAGG CTAGATGAT ACAAAGTACC AAAGTTGCTT  
1440 CAGCTTGCT ATGAGTGTG GGCATTAATC TCTAAACATT CCMAAGACC ATGAGCTGAA  
1500 CCTAACTCC CTGCGAATC TGGACAAAG GAATATGAAA ATTCCATTT GAAATCTGAC  
1560 CAGCTATCT GGACTCAGA CATAGATCAG CCACTGCCC AAGCCATTT CAACTACAGA  
1620 AATTAAGAG ACTACAGCTA AATAAATTTG AACATTAAT AATAATTTAC CACTTTTGT  
1680 CTTTAAAGC ATATTTGTAA ACTAGAACT GAGCAAGAT GACTTTACTT TCTCAAGTTT  
1740 GATATGATG TGACTGTGTC CTTATCCCTC ACCCTTCCC TTCCCTTCC TAAAGCAATA  
1800 GTGCACACT TAGGTATTTT TTCTTCCGA ATTGTGAAGA AAACTTAAT GCAATGATT  
1860 TTTTCTTTT GCAGAGACCC TGTATTATCAT CTGTTTAAA TGTAAATGTC CCTTATGCT  
1906 TTTGAATATA ATTCTCTTTT GTAAAAAAA AAAAAAAAAA AAAAA

(2) INFORMATION FOR SEQ ID NO: 138:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1935 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

60 TCTGAATATA TGTATACAGA TCCCCCTGAG GATATCTTGA TGGGCTGAGC AGCTGGCTGG  
120 AGCTAGTACT GACTGACATT CATTTGTATG AGGCAAGCTT TCTGTACAG GATTTTACCC  
180 TCTATGTTTT ATATACATT TCACTGTGAC TTGACCTCA CTTTACACAA GAGGAATATA

	TCGAAAGTGA GCTGATGCC TGAAGTCAC TTGAGTAACT TGGCAAGTCC ATGCTTCCA	240
	CTGAGCTGCT GAGGTACAGA AGTCTACTTC TGTGCTATT TTGTATATC TCTTATAT	300
5	GTGCTGACT TTGAAAGTC TGAATGGGT CCGTGTGTC TTTTACTTT GAAAGAAATCA	360
	GTTTTGCTT CTTTGTGAA AAGAAACAA AGTGCATTG TTTTCTACT GAAAGTACG	420
	CAATGACAG AGGTGAACAG GAGCTAGTT AGGCTTCTT GTAAACAGAA AATCATATCA	480
10	AAACATATC TTGCAATG TTTCATATG CTTGCTACTT CTTGTAGATA TTTCATTTCA	540
	GGAGACAC AGTTAAAGCC GTGAAATTTG TACTTAAAGAA CTTGGATCCA AACCTCTTC	600
15	CACATATTG CTATGCTCT GAGCAAGTT TTTTTTTTT TTTTTTTTA ACCCTTCTG	660
	AGCTTGACT TTCTATGCTT AACTCAAGA ATTGTGTGA GAGTGAAGT AATGCATTG	720
	TAAAGGCTT GCGAGATAG AGAGTCTAG TTATGATTTT ACAAAGTTGT TAAAGCTGTA	780
20	AGAGCTTAAA AACTGACATG AATGCAATG CATTTAGCCC CACTGACTG GAGATATGTC	840
	AAACATGACC AGAGTCTCT TTTTCAAAAT ACTTACAGT TATTCATAT AAAATTTTTC	900
25	TAAAGATTA TCTTATTTAT CTAACTTAAA GCTTCTGTTT TATTCAGCT CCGTATATC	960
	TGGATATAGA TAAATGACCA GATTAAGCTTA ATTCTAGCT GGTGCTGCTT GATGTATAT	1020
	TGTAGATAG GACATTTCT TTTTTCACG AGCTGTAGT GTGAGAGACC TGTGGAAGG	1080
30	GAGGGGGTGA GATGACACC ACGAGGGGAC TGAAGTGGCA AACTGAGT TCTTTTACT	1140
	GTGGGTATGT GAGCTGCTT TCTGTATGTC GTTTGCTTAG GAAATGTCT GTTCCCTGTC	1200
35	TTGCTATTTG CAGTACATTA CAAGTGGACC AACCCAGTTA GAGTATCTT AATATATATC	1260
	AGTGTGTCG CTGACTCTCA ATGAGCCGAC CAAAGTGGC TAAAGTTTGA CAGATACAT	1320
	AATTAAGCAA CCTAAATAT TGAATGCTGT GTTGTGTGA CATATGCTA TGGCAGACAC	1380
40	TGAAGCTTAG AGTTATATAT CATGTATTAG GTTTCTCAG AAGGACAGAA TTAGTAAAT	1440
	ATATGTATAT ATGAAAGGGA GATTATTTAG GAGAACTGAC TCCCAAGTT AAGAGGCGAA	1500
45	GTTCAGCAAT AGGCTGCTG CAAGCTGCTT TGAAGAGAG CAGATATGTC CTGAGCTGTA	1560
	GTTCAGAAAC CTGAAACATG GAGGAGCTGA CAGTGCACG AACCTTCACT CTGTGCGCA	1620
50	AAGCCAGAGC CCGCTGCGAA CCAAGCCACT GGTGCAATC CTAAATTTCA AAGCTGAGG	1680
	AAGCTGAGT CTGATGTGCA AAGAGAGGAA GAGTGAAGAA AAGCCAGAGC ACTGAGCAA	1740
	CAAGTGAAC AGTGTATAC AACAAATGTC GATTAACAAA GAGCTTACCG ATTCTTCTCT	1800
55	GCTTACTGTA TCCCTGAAT TCCCTGTCG TGTGCACTT CTAAAGCTAG TTCTTAAAGG	1860
	CTTTCATTTA CATGAGCTGT CTGAAAGCCC TGTCAATATAT TGTGATGTA AGTTTCAAAA	1920
60	AAAAAAAA AAAA	1935

5	(2) INFORMATION FOR SEQ ID NO: 139:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1446 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
10	(D) TOPOLOGY: linear	
	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 139:	
15	NOCCCTCTG GACAAATCA GATTAAGAC GTTCTGCGA GAGAGCTCTC AATCTCAGA	60
	GAGACAGAC AGCAATTTCC TGTGAGGAGA GAGAGAGATC CAGCAATCTT GATCTGCTCT	120
	GAAACTTAT TTTCCTCTG CAGAGATCA TTTCCTGAG TGGAAACAG TTCTGCAATG	180
20	TGATGTGTC TTTCCTCAG CAGAGGCCC CTCTTTCCC AGCACTTCCC TGCCTGCCCC	240
	AGGCTCAGG CAGACAGCA GTTCTCTCTC ACATGACAG TGAACACAGA CTTCATATTTG	300
	GAGAGACTG AAGAGGGTTGA ACAAACCTG AAGGAGGCCC GAGCTTCTCT CCGAATTTG	360
25	GGAGAGGGG TGTGACAGC TGCCCCCCC AGAGGCGAG CATGTTTAC CAAGGCTCTC	420
	ATTGTGCTG GAGACAGACC TTTTCCCGAG CCGTCAGAC ATTGCTATC GGTGCGAAAC	480
30	TGGATATAG GATTGAGAG GAAAGATTC CAATATGTC CAGAAATTTT CCGTCTCAG	540
	GAGAGGCGAG GAAAGCTAG GCGAGCAGG ATACAGGGG AGGAGATGTC CAGGTGAGGG	600
	GCTTCCCGCC TGTGCGCTTT CTGTCACAA TGTCTCCCC AACCTGCTTC AGTTCTCCCT	660
35	TCCCCCTCAT CTGCTTCCC CTCTTTGAG CTGTCCCAT CTGAGTGTCA GACAGAGCTT	720
	CTCTTCAACT GACAGACTTC CTGTGACCA GCCCCCTCC TTGTCTGAAA AAGGAGGCTT	780
40	TGAATGTGTA AGGAGGCGAG GTGTGACAG GACAGGTGTC GAGATGAGG	840
	TGAGGAGTGC TGGGAGACAG ATGAGAGGGG CAGTGGGAGC AGGCTTTGGG CAGACACAG	900
	CAGAAATAT TTGAATGTG TGAAGTGAAT CCGCGAGAGC CTGGGCTTTG GCAATTTGGG	960
45	AAAGAAATGA TGTGTGAGG GACTTAAAGG AACAAGTGA CAGAGGAGAA GTCTCATAT	1020
	GCTGCAATTT TGTGGGTGT TATGTCCAAA CTTAATATG GAGTGGGTCG CTGTCTTCA	1080
50	CTGACAGCA AATTCAGAT CCGTGTCTT GATGTCCGAG AACTTTCCCT CTTGAGTGT	1140
	CGTCTCTTC CTATCTTCAT CCAATGAAA TTATGTATTT TCTGATCTTT TCCCTGCTCT	1200
55	GATCTGATC CTGTCCAAAC AGCCAGTCCC TCCAAATCT AAGAGATGTC GCGCTGAAC	1260
	CTTGACAG AGTCCCTCAG AATTGGAGCA TGGGAGGGGG GTTGGAGGAC CCGATGATTC	1320
	AGCCAGGAGC TGCATATCCC AGCTCTCTTC CCGAAACAAA TCCCAAGCAT CCGTTATCTC	1380
60	TACCCCAAC CTTGGGCTCT CTGTACAGAT TTTTAACTT GCGAAAGAT GAGAGATATA	1440

TTGTAA

1446

5

(2) INFORMATION FOR SEQ ID NO: 140:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1109 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

TTTTTTTTTT TTGATATGA AATTGCTTTT CTCATTGCA GAATTAAGCT AGGGAACAC 60  
TAACCCAAA ACTTCTGTA GAGCTGTTC TTGGAAGCA GCATCACTTA TTGGAGTAA 120  
AGACTCGTA TAAGGACAC AGCTGCTTA CTGGGTGAT GGGGTTAAT TTTATAGAT 180  
TCCATTTTCC TAGTCCACA TGTGAATG GATTTTGTG ATCTTAATCT ATATCTAC 240  
CTTATATAA AGATCAAAA GATTAATCT CTATGAACAG ATTGGAGATA GGAATGAAA 300  
AGTTGGAGG ATGCTTTAT TCTAATGTA GGTAGCGAA ATGTGATA ACNTACTGG 360  
GCTGAGGAG GCATGTTCT TTAGTTGAG TTCTCATTTT TATCTCCAG TACTGACTTG 420  
TGGGAAAGC ATACTTTTC ACTGCCAGT ACTGAATGCA GAGGCTCAGT GAATATATA 480  
TGTGGAGT GCATGCTTT GTTTATTAG CAACATAGC TGTATTAGA CAAGTTGTT 540  
GTTTTGAAA GGGTTAAG CTTAAGTGA ACAATCTAG CTACAGTGA ATGAAGTAG 600  
TAATATACT TGCATATTT TAATTTCCTT TGGTTAAGG TCCCCATAC TTCTCTGTT 660  
GGAGACATA GAATATGAT TACTTCAGT TTAGTTTCT TATTTTTTT TTTCCTCAT 720  
TTGTCCCTTG TCATTTGTT GCACTGTA BATCTGTGG TTATACATAG GCGAGCTCT 780  
TGTGAAGTG GTTATTTCA CTGAGAAAG GGGATTGAAA ATCAGTTAGA ACCAATGAT 840  
TTCTTGCCC ACGGAACAT ATTCTATDA GATAGCTGA AAGAGTCTGT GTGAGAGCT 900  
CAGCTCAAA CACAGATCA GCACCTGTA TAGGAATTC CATGAATAT GACTTCTAT 960  
TCTGTTTTAT CAGATGAT ATATGCTTA CTTCAGAAA ACTAAACAG TCATTTACA 1020  
AAGAAAGCA ATCTGTATCC TAGCATTTT ATTAATAAGT TAAACAAA AATTAAGG 1080  
GACTCGAG,GGGGGGCCCG AACCCAT 1109

55

(2) INFORMATION FOR SEQ ID NO: 141:

(1) SEQUENCE CHARACTERISTICS:

60

(A) LENGTH: 497 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

TAGACTAAC TTAAATCTT TTATCATCT TTATTTATT AAAAAATTT ATTCTTGA 60  
10 ATTTCCTGT ATTTCCCTTA RCTCTCTA TAATGCTTA TATTCATGT AACCAACT  
CATATCCTT AACATTAAT CTCAAAAGC TTTTATTTT TATTTTITG AAGGTAGTTT 120  
TTCTGTGT ACTCTGTA ACATTTTTC TTCAATCA TTGTGTGCC CCATACAAA 180  
15 ATGCTTTTA TTTTGAGA TGTGACTT TTTAGTAGG CATGAGTGT CTAAAGCCA  
GATATCTTC CACATCACT GTTGCTTG ACACCTAGT TTTAATCTC CATCTTACT 240  
20 TTAAACCTG ACATGCAAT CCTCAGTCAG GGCAGAGAC GGCCTGAGC OCTTTGTGA  
GATGTCAC CACAGCAGA AGGCTGAGC CTGGTACTT GTACCTGTT ACTTGTATA 280  
AAAGATTA TCTAAA 497

(2) INFORMATION FOR SEQ ID NO: 142:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 269 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

ATGAGGAGA GCGAGCTGC CTCGCAACC CTCCTCTCA GGAATGGCT TCCCGAGAA 60  
40 TCCCGACAC ACATACCTC TCTTTTCTT CTAGTCAAC TCTGTATTAT TCCTTGCCT  
GCCTCCCTC TTCTCTCCC TCTCAACTT TTACTTCTG TTCTATTTT ATGGGATTTG 120  
45 GGTGTAGT TAACCTACA ACAGTGGCC CACACCAAG TCTTCGAGA AAAAAATCA  
AAGATTTA ACAAAAAA AAAAAAA 269

50

(2) INFORMATION FOR SEQ ID NO: 143:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1269 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

60

TTGATGTAAGT ATGATCTCTC CCGATACAG GAAAGATCTC CCGAATGAAA GCGCATGAC 60  
TTTATACAGT CCAAGACAT CTGCCCCG CTGTGAGCT GCGTGGAGCT GTTGGAGCTC 120  
TTCCGAGCTC TCGATGAGGT GCGCGGAGAG GCGTAAGCTC GAAATGCTGT GGTGTGATTC 180  
AAGCGGACCA CCTGAGGAGT GCGCAAAAGA TGTGAAAAG TCTTCTATGC TCGCGGTGCT 240  
AAATGCTGTC TCTGTGAGCG GAAATGTGAG GCGCTAGAG AGCTATCAG AAGATTCAGC 300  
GCTTCATG CCGACAAAGT GCGAGACAC AAGCTTACT TGTGAGAGT GAGCTTCACA 360  
GACTGAGAG CCAATGTTGC AGCAAGAGT GAGATCTGTC AGTGTCTTGG GTATGTGAC 420  
ATACTGTCA ACAAATGCTG GATCAGCTAC CTGTGTACA TCAATGACAC GACATGTGAT 480  
GTGACACAGA GGTTCATGCA GACAACTAC TTGTGCCCCAG TTGCTCTTAC GAAAGACATC 540  
CTGCGCTCCA TGCATAGAG GAGCGAAGCG CACATTTGCG CCAATGACAG CATCGAAGCG 600  
AAGATGACCA TTGCTTTTGC ATCGACATAT GCAAGCTTCA AGCAGCGAC CCAAGCTTTC 660  
TTTGACTGTC TCGTGTGCGA GATGGAACAG TATGAAATGG AGGTAGAGCT CATCGAGCCC 720  
GCGTAATCC AAGACAAAGT CTCTGTAAAT GCGATCAGCG CCAATGTGATC TAAAGTATGA 780  
GTATATGACA CCAAGACAGC CCGAGGCGCA AGCGCTGTGAG AGGTGAGCCA GAAATGCTTT 840  
GCTGTGCTGG GGAAGAGAAA GAAAGATGTC ATCTGTGCTG ACTTACTGCC TTGCTTGAAT 900  
GTATATCTTC GAACTGTGAC TCGTGGGCTC TTCTTCAAGC TCAATGCTCC AAGGCGAAGA 960  
AAGAGCGAAA ATTCAGAAC TCTTATGACT CTGACAGAGC AAGGCGAAGG CAGAGAACCA 1020  
GCACTCTTAG GCTTGCCTTAC TCTTACAGGG ACGATTCAT TTGTGAGAC TTATATGAGAG 1080  
ATTGTCTCA GAAATGAGAA AAGCTGAGA AAGACATCTC GTTCAAGATC GCTGCGAGAG 1140  
GACATTCAAA AAGCAGACCA AGCTTCTCC CAGGGTGAAG GAAACACTT AAGCAATGAAA 1200  
TATGAGAGCT GGTTTTACA GTTAAACTTA GAAATTAACA TCTCAACAG TAAAAAAAAA 1260  
AAAAAAAC 1269

(2) INFORMATION FOR SEQ ID NO: 144:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1944 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

AAAAAGCAAA GTATAGATA AAGAGAGCC CTTTGTGAAA ATTAATGTGC ATTAGAGTGT 60

TTTACGCTGT AGCTGTTTAA GTTAAAGAT AAGATATGCT GCGTACCCAC CTCAAAATGT 120  
CTGTACTGCA AAGAGGAGCTT GAGCTCTGAC TTTCATTAAT CAGCTTTGAC CAGATGTGTA 180  
GTCCCAAGAA AAGACATGAG TGCAGATGAG TAAAGAAATGG AACTGAGCTTG TGCATATGAGC 240  
ATGAGACACA AAGGATTCACA GCAATGCTCC TGCCTTACAG TGGCAATGAG GAGACATGCC 300  
AAGACATGAT GTTCTTGCCA CCGGATGTGG TTGTCTGAG TGGATCTGCA TGTCTGTGCC 360  
TGACCTTTAT TCTTGAAGCT GAGGTTTAC TGCATCTGAC TACGTAGAGCT AAGAGCCACA 420  
GAGAAATGAG GTTGGAGCTG TGGCCCCGAA ACTAGGAGGAT GTGGATTCAT CACAGTGTGG 480  
CCTTTTGCT CATTAAAGATA GAGATCTACT TTGTAAGGGA ATTGTCTCTC CCAATTAATAT 540  
TTGCTTTAC TTGATCTTTT CTTTTGTGCC AGTATTCAG TGGTATTAAGT CTGACAGAGG 600  
TGACATTTGG CCAAACTGAA CAGTGTCTTG CTGATCTTC CTTTGCGAAA CATCGAGGTC 660  
AAGATTCAG ATAGGCTTTC CTAGGCAACT GACATTTCTG GCAATGGAGCG TGTGTTTCCA 720  
CAAGTATTT TCAATTTTAC TGCAGATGT CCAAGAGCTG CTGTGAGAGT GAGGTGTGTT 780  
CCCCCTGCC TGTGTGAGC TGTGAGAGG ATTCAGATCT AAGGATCTGAG AATTCAGAGC 840  
CAGAGACTC CTGTGTTTGT GGTCCAGGAT CCTGTACTAA GAGAGTCTTG GCGAAGGAAA 900  
CAGACAGCT TTTCAGCAAT GAGCGGAG GAAACAGTGG GTTGTGCTGG TGTCTTACTT 960  
GTCTGTAGAC TGTGCTGTGG GCGCAATGAA AAGTTAGATC TGTGATCTCT GGGGTTTTGG 1020  
TGCATTTGTT CAATGCTTCC ACTGTAGAGC AAGCAGACA GTCTATACTC TCCCAAGCTT 1080  
GCTGTACTC CAGGTAGAC TGCATAGAG ATCTGTGAAT ATTGTGATGG AATATCTTGG 1140  
GTATGTAC AGTTGAGCTG CAAATCACA AATTCGAGAG TACATGTGAA GCTATTAAGAC 1200  
AGTCAATCTT AAGAGAGAG GATAGCTGAG ACTCAATGGA TCAATGAGTTC CATGCTGAGT 1260  
TGCAGTTC CTTTTCTT CTTCAGGTTT TGTCTTTGCC TGTGTGTGCC CCAAGCAAGG 1320  
AAGACATG GGTGTGATGG GAGAGACAGA TTAAAGATAT ACGAATGAAA CCGAAGATGG 1380  
AAGATGAGG AGCTTACTGT GAAATGAGAG AGTACTGCT GCAAGAGCTG GAGGTCAAGT 1440  
GTGAATGCTG TATGTGACAA GGAATTAAT ATCTGAGGT CTGAGAGCTT CAGCTTGTCCG 1500  
TCAAGTCTT CCTGTATAC TGCATGAGCA CAGATTCGAA GTTGTAGCTC TGCACCTTAA 1560  
ATCACAGCT GGGCATTTGC TGGGCTCAG GCTGTGAG TGTCTTACT GTTTCAGCT 1620  
GTGAGCTGAG ATGCTGAGAC TGAAGGTTTT TCTGTGCTCA GACTGTAGCC TGTAGCTCTT 1680  
GCGCTGTGTA GAGCCCCCTC CTGTGCCCCC AGTGTGCTGC GTTGTATAC ATCAATCAGAA 1740  
AAGATGAAA GGTTCAGAGG AATTTTCTG CCGTACAGAG GTTGTAGAG AAGAGACACA 1800  
GTATTTCAAT GAAATTAACA TATATCTTGG TTTTTCTTCA ACGAAGAGT TATATGAGC 1860

ANTGTCACT GCTCAAGTT GAGTGTTA TTCAATATA ACTGTAAATT TCTGATATA 1920  
AAAAAAAAA AAAAAAAAAA AAG 1944

5

## (2) INFORMATION FOR SEQ ID NO: 145:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1021 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

TCGACCCACG COTCCGGGT GCGACGGG GAGTTCGGG TCGAGCCGG TGCCTCGGG 60  
CGCGCCCTTC ACCATGGCT CGCGAGCT GACTACAC ATCGAGTCC CGGATCAGC 120  
CTCTCGAGC CAGAGAAC GCGCCGCC AGGTGGAG GAGCGAAG CTGGCAGCC 180  
TGTGTGATT CTYTGAGT GGGTGGTG CAGGACAG AACTTTCCA AGTACAGTCC 240  
CATCTACAC AAGAGGGCT GCATGTAT CGGATACAC GCGCCGTGG ACATGCTTT 300  
CTTCTCCGAG TCACTGGTA TCCCTTCACT TCGTGTGTTG GCGCAGAGC TCGTCAGCT 360  
GCTCTTTGAT TATGAGATG AGAAGAGGC CTGCTCTTC CATGTCTTCA GCAAGGTGG 420  
CGTCATGCTG TACCGTAGT TCTGAGCT CTGCGAGCC GTGTGCTTCT GCGCCCTGG 480  
TGTGTGGGC ACCATCTTGG ACAGCGCTCC TGTGTACAGC AACTGTGTAG GGGCTCTGG 540  
GGCCCTGGCA GCGATCTTGG AGCGCGGGC CGCATCTGTG CGCTGTGTC TGTGCTGGC 600  
CTTTCCTGCA GTGTGCTGCC TGTTCAGCT CTGCTTGTCT CCATTCACAG CCGTCTTCCA 660  
CACCCAGTTC TATGACAGGC TACAGGAGGC GGGCTCTGCG TGGCCCGAGC TGTACTCTTA 720  
CTGAGGGCT GACGAAGTAG TCTTGGCCAG AGCATATGAA CGCATGGTGG AGGCAAGCTT 780  
GGCAGCGCGG GTCTGGGGC GTTCTGTGTA TTTGTGTGTA TCTGACAGC TCGCCACTT 840  
CGTCACTAC CTAATTAAT ACACAGGCT CTGTGTGAC TTAGTGCCA ACTGGTGG 900  
CTGTGAGGC CATGTCTCA TCTCACTCT CTCTCAGAAA TAAATGCGTG ACACCTGCC 960  
50 AAAAAAAAAA AAAAAAAAAA ACTCGAGGG GGGCCCGGTA CCGATTCCG CATTATNAGG 1020  
T 1021

55

## (2) INFORMATION FOR SEQ ID NO: 146:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1285 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

GGCAGGGA GGGCAGGC AGCATGGG CTYTCAGTT CGGTCTCTG GTGTAGGGC 60  
AAGCCAAAT AGGGATTGC GTTCCCTCCA GTCCAGACC CTATCAGAT TGGATATGC 120  
CTTCATATT GATTGANTT ACAGTGTGTT CAGAGTGTG CTACAGTTTT TAGGATATA 180  
TAGAATACT GTTAACTGG TATTCTCTG ATTGTATAT GAGAGAAAA CAACTTGTCT 240  
ACATGCTA AAGATGACA GACTTGGACA ACATGTCCA ACATTACATC CCACTTCCA 300  
AGAACTGACC ATTCTGGCA TGAAGTTTAC AACTTTGAT CTGGGTGGAC ATGTTCAAGC 360  
TCGAGAGTG TGGAAAACT ACCTCTCTGC TATCATAGGC ATTGTATTTC TGTGTGATG 420  
TCAGAGCCAC GAAAGGCTGT TAGAGTCAA AGAAGAACTT GATTCACTTA TGCAGATGA 480  
AACCATTGCT AATGTGCTTA TACTGATTTCT TGGGATPAG ATCGACAGC CTGAGCCAT 540  
CAGTGAAG AGGTTGGAG AGATGTTTTG TTTATATGCT CAGACACAG GAAAGGGAG 600  
TATATCTCTG AAGAACTGA ATGCCCGACC CTYAGAGTT TTCATGTGTA GTGTGCTCA 660  
AAGACAGGT TACGAGAGG GCTTCCGCTG GATGGCAGC TACATTGAT AACAANAAT 720  
CAATGTGTT CGAGTCTCA AGTGTAGGC TTACTCAGAG ATTGTATGCT TCAACATGCA 780  
TNACTGAT TCAATAGACT TTTCTGZTT ATAAACAGA TGTTTTTTAG ATTATATA 840  
TTAATCAAC TTAATTGNA TGAATATGA AACTGATTC AAGTAAATTT GAGTATCA 900  
ATGTTAGTT TCTAHTCCA TAAAGTACT TGTTTTTTAC AGTTTATAT CTGACATC 960  
CCAGGCCA TTTGTNAGA GCACTTTCC ACAGTACAT TTGAGCACT TTTTACAC 1020  
ATGAACATAT AAACATAT TAAAGCTCA TCAATTTAA TTTTTTATCT ACTTTCTGG 1080  
AACTAGTTTT TAAATTTAG ATTATATGTC CACTATCTCT AAGTGTACAG TTAATATA 1140  
GCTATTCAA TCAATGATG ATGCTTACA GTTTGATA ACTTTTTTCT TTATCGAAC 1200  
GTCACTCAAT AACAANAAT CTATGTTTG GCAAAAAA AAAAAA NTCAGGGG 1260  
GGCCGTACC CAATTCGCC TAAAG 1285

T

## (2) INFORMATION FOR SEQ ID NO: 147:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1386 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

60

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5 GCGACGAGGT GCGCGAGGCG TCGATGGTTC TCTGGGCTCT CGGACGAGGT GAGCAGCTTC 60  
ATGAGAGCCA GCGTCTCGAT GCGGACCTTC GCGGAGTGTG GAGAGTGTG GCGCCCTTCC 120  
CGAAGGCGCG CGAGAGCGCG TTTCAGTTAT TTACATGAGC TTGAGCAGGT 180  
10 TTTCATATGA TCGAAGCCA TCGCCTCTTT CTTCAGATAT TCTGATATAT AGCAAGATCA 240  
TCAGATGAGA GTGTGAGAAA GAGCTGAGAG CGCTCTTTCA CGAGTATTC CGAATTGAGA 300  
TCGACCCACA CGGACCGTTC AAGAGAGAG TACCTCATAT GGTGATATG TCGACAGAG 360  
15 CGCAAGATCT CGTATGTGAG CAGAGATTC AGAAGTTTCA GATTAGCCGAG GTGGTTGAG 420  
AGTCCATATC AATCTGTAGG GAGGAGATTA AGACCTTCTT CAGACAGCTC TTACATATGA 480  
20 AGATCCCCCT TTTCATCTTT TCTGCGGCGA TTGTGATATAT CTGTGAGAGA ATTAATCCAG 540  
AGATTGAAAT GTTCCAGCCC AACATCCACA TCGTGTCTAA CTACATGAT TTTAATGAG 600  
ATGGTTTCTT CGAGGATTT AAGGCGCAGC TGAATACAGC ATACAGAGAG AACAGCTCTG 660  
25 TGTGTAGAGA CTGTGTATTC TTCCAGAGAC TTGAGGCGAA AACAGATATC ATTCCTGAG 720  
GAGACTATAT CGGGAGCTTC ACCATGCGCG ATGGGATTTCC TGGTGTGAG AACATCTTCA 780  
30 AATTGGGCTT CCTGATATGAC AAGGTGAGAG AGCGGCGGGA AGCTACATG GACTCTTATG 840  
AATATGCTCT GAGAGAGAGAC GAGACTGTG ATGTGTCTAA CGGCTATCTG CAGCAGATTC 900  
TGTCCAGGCG GGTCCAGCTG GAGATGTGAG GCGCCGTGAG GCGCAGGCTT CGAGGCGGCG 960  
35 TCCAGGCGCT GGTGAGAGAG GCGCGCTGCC CAGATTTTCC TCCCTCTTGA ACAGAGACA 1020  
GAGGCCAGCG TCGCCAGCAG TGGCTGAGTC CTGTGCGGCC CGTCTGTCTT CTTTCCCTG 1080  
40 AACAGCTTCA TCAACAGAGG CTGAAAGAAA CGCGCGCATG TCCAGAGCCA CAGCAGACT 1140  
TCTGTGTGAA CTTTGTGACA CAGCATGTCA GTCTCTTACG GATTGTCTAC TCCAGGATTT 1200  
TTCTTCAAAA TTTTATTAACA TCGAAGATTC AACAAATAT AATGTGTGAA AACATGAAA 1260  
45 AATTTTAAAA TGAATAAAAA GCTGCTTGA TTCCAGGGAAT GTGGGTGCGG GTTGAAGCTG 1320  
GAGCTTTGCG CTGTGCGGCA CAGGAGATCG TTCTAGAGAC AACATTTGAG AACAGCAAA 1380  
50 AAAAAA 1386

(2) INFORMATION FOR SEQ ID NO: 148:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2098 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

5 AGCCCTTCTC CGCGCGCTTG GAGCTGTGAC ATCTTAAAGCC TCCAGGCTGCG TGTCTTTTTC 60  
TGGGTGAGCC CATTGTCTGTG ATCCAGGATT CTGTGAGCTG AACAGAGAAA GCGCTGTGAA 120  
CGCTTAATGCG GTTATATATC CTCTACAGCA GCGCTTACAC CCGTGAAGATG GCAATTAATC 180  
10 CTGGCAATGCG TTGTGAGACA TGGGACAGAC AGTGAAGGCC ACCCACTGT GTTATCTGCG 240  
CTTCTATATG AGTGAAGACC TTGGGAGAAA CGGCGCTGTG ATGAGAGAT GTACTGTGAGC 300  
15 CTGGGCTTAC CGACTGGGCT GGAATCTCT GGTATCATGTG ACTGTATATC CAGAGAGAGA 360  
AAGTGTATAT CTGAGGATTC AGATTCAGC CAGCAGCAAA GATTATATTT CAGAGAGACA 420  
GAGCTTGAAG TGGGCGCTTTC TGGCAATTTCA GAGTCAAGCA AAGCAGCAG GCGCAGAGAG 480  
20 CTTCGTGCGG TCTACAGAG AAGGTTCCTG TGAAGGCTAT CAGTTTGTTC CTTCAGACTT 540  
GCTGTATATC TTGGGCGCTTC GCGAAGCC TCCAGAGTC CGCTGGGCTGT GATGGCAATTC 600  
25 TGTGCAATCC TGCCTTTTTC CGAGCTTCTG CAGAGAGGCC TCCCAAGCA ACTTTTCTGCG 660  
GCGCTTCCCT GTCCAGCTGCG CTGATATCAT GTTCAAGAAA CTGAGACTGCG AAGGCGAGCA 720  
CTTCTTTCCA CGTTAGAGATG AAGTCTTCCG CGCCTGTCTC TGGCAATAAA AACACTTTAA 780  
30 AGATGAGACC ATGTGCTTCT TTGCCCTTCT CTGTCTGTAT CCAATCTTCT GCGTCCAGAT 840  
CACTTCTTGG GAGACTATGCG ATTCAGTGTCC CGCCAGCTGT GTGGGAGAGC CATTGTCTCT 900  
35 GTCAATCCAG AACCTGCTCT GAGGTGTGAG GTTACAGAG ATTCAGAGCA GCGTGAATCC 960  
TCCAGGCCCA CTGTGCTGAC CGTTCCCTTC CGCTGCTCTT TCTTTTCA CAGCAGACT 1020  
40 AACCTTTTCC ATTCAGATTC AACATAGGCC CGTCTCTTGT GAGCTGAGAC CTTCATTTTA 1080  
CGTGGCTGAG ACTGTGAGCC TCTGTGAGGA GTGGGATTTCA CTTTCTCTT GCGCTATAGCA 1140  
GTGCAAGCTT CAGTTCTGAC CGAGCAGAT TGAATCATCT GCTTCCATGT CTCTGGGCT 1200  
45 TTGCTGTGCG CGTGAAGACT AGCTGGGCTG GTCTGTGCTC CAGCTTCTCT CGCCCTCTCT 1260  
GGAATGTCCCT TTGGAAGGCC CTGTGTATCC TCCGAGACA GTTGTCTGCG CTGCAATGCG 1320  
AAGTCAATCA GGGGCTTTTA CGCTGTGAC CAGAGATGT AACAGCTCC TGCATTTAT 1380  
50 CTTCATGTG TCACTGAGCG TCCAGAGCT TCAATGCCCA TATTTCCTG GCGCAGAGAG 1440  
GCTGAAGTTC CGTTTCTGCG GATTGTGAT GAGCTGTGAT AACAGCTTAC ACCCGAGTGG 1500  
55 CTCTTCCAT GATGTGTGCT GCTCATGTG CTATGCGAA CTAGAGACT AGCTGTGATC 1560  
TGGAGATGCG TTTCGCTGCA GAGGTCTTTT GCTGTCTCT AGCTGTGTC ACTGTGTCTC 1620  
60 CTGCGCGGCT GTTATATATC CAGGCGTGT TGCAGCTCTC CGAGTATCT TGGGATCTCA 1680

1740 GCTGGAAATTC CTTTGTGGTT TGTCTCTCTG CTTCTCACTC TTGTATTATAG AAGGATTTCA  
1800 CAAGGGGAGA GTGGCANTCC TGTCTCTCTT GTGCCAGACC AGAGTTTTCCT GAGGGGGCCT  
1860 GACCCTAACC CTCAGCTCA GCCCTGTACA CTTGACCCCTG TAAATGAGTG GGGTTTCTCG  
1920 ACTGTAAATCC CTGACACAG TAAACACAA AGGACTCTTG GGGGCTCACT GTGAGAGCCA  
1980 GGGTACTCTA CTCCTCCAG TGAGACAAA CTCCTAGGCT GTATCCCATTA ATTTCAGAT  
2040 GAGAAACNTT AACNTAATA ATTGTAGTA AACATAACT CATGAGACT AAAAAAATA  
2098 AAAAACTVGG GGGGGGGCCC GTAAACCATT GGGCCCTTNG GGGGGGNTT TTAAATTT

15

## (2) INFORMATION FOR SEQ ID NO: 149:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1847 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

60 TCGACCCACG CTTCCGAAC TGGCGGCG CGGAGCGCG TTGKGTCTG GTCTTCGCT  
120 CGGCCCCGCG GACAGAGCG TGGCCCCGCG GCGGGGAGAA GATGTGTGCK AGCGGCTCG  
180 GCGCGCCGAC GCGCGCGGAC GATGTAGGCC AGCCCGACCG CGGGGTTCG CCGACAGCT  
240 GCGCGCGCTG GCGCGCGGCG GCGGANTGCC GCGCGGCGCG GGGTGGAGCT GATCAGAATA  
300 ATGTTTACGA TCACCCCTCT GCGAGAGCTG AAGGTGTACA TCAGAGTGT GCGTCCCTCG  
360 GTGCTCTTCA TGATCAGGCT AAGGCCATG GCGATAGCTT TCTTGACCT GGGCTACTTC  
420 TTCAAATACA AGGAGATTA ATCCCCAGAA ATGCCAGAG ATTGGAAATC TTCTCTCTA  
480 CGGTTCAATG ATTGTGACTT GTGTGTATCA GAGAAATGAA CCTCAAGCA TCTCAAAAC  
540 GACACACAA CTCGGGAAG TACATATACC AGCGGGAGG CCGAGCTTC CACCCGCTC  
600 CCGCAGGCC TGGAGGACT GGGCGCGGTG ATATCTTCAG TCTCAATCAC CTTAACCTG  
660 GACCCACTGA AACCTTCTCG AGGATATTC GCGAGCTCA CCTATCTTA CTCAACATC  
720 TTAGGGGATC AGATGTGACT TTCAAGGAGG GAGGCCGAG AGGAGATMA CATCACTTC  
780 ACCCTGCTTA CAGGTGGAG CTCAGATGAC TGGCGCTCC AGGTCAGTG TGAGCAGTG  
840 GTATTCACAG CTTGCATGAC CTTCAAGGCC AGCCCTCGCG TGTTCCTCGT CACTGTACAG  
900 CCACCGCACT GTGTCTCTGA CAGTAAAGC AAGGCCAGCG TCTGTACAA GATCTTACA  
960 ACTGCGAGAG ATGCCAAGAC AATATAGCC CAGATTTACA ATCTTTCTG GTTTATAG  
1020 GGGGCCATG GAAAGTCTA TCATCTTTA AATCCCAAGC TTACAGTGT TGTTCAGAT

60

1080 GATGACGCTT CATTAATAA TTGCACTCT ATGCACACA GTTACTTCT CTTTGTGATG  
1140 GTGATAACAA TGTTTTCTA TGTCTTATC AAGGGAGAC CTAGCAAAAT GGTGTAGAGC  
1200 AATCTGAAAT TTGTGCCGA GAGGTGGCT TTGGCTGAG CTTAATTOCA CAGCTCTCTG  
1260 TTTTTTCAGA GAGACTGAGA GAACCAATAT CTTTGGCTGC TGAACCCAGC CTGGCCCTCG  
1320 ATGCTCTGTG AATACATAT CTTTGGATGT TGGGTTATTC CAGCCAAAGA CATTTCAAGT  
1380 GCGTGAACCT GATTTGTACA TTTTATATA AATCTATTTCA GAATTTGTC CATTAATGCA  
1440 CTTGCTTTCC CTTGGGTACA GCCAGAGGCC TTCAACCCCA CTTTGGACTT GAGGACCTAC  
1500 CTGATGGGAC GTTTCAGCT GTCTTAGAG AAGGATTCCT GATCTAGCT GGTCAAGAGC  
1560 ATGTTTTCAC CAGGTGACA GGAGCTTGC GTGGCTGATG GGGTTCAGCT TTGCTTTGCT  
1620 TCTTGTTTCA GCGCAATATG TAGAGACAT TTCAAAACAGT CTGCACCTTT GATACGGTAT  
1680 TGCATTTTCA AAGCCACCA TCCATTTTGT GATTTTATG TGTCTGTGCG TTAATATACA  
1740 TAGTAAACAC AATATATCTT TTTTCTCAT TTTGCTTCCA GGAACATAC CTTAAGTTTT  
1800 TTTTGTTTTG TTTTGTTTT TTTTGTTTT GTTTTCTTT ATGAGAGAAA AATAAATAG  
1847 TCACATTTTA ATACTACCA AATATGACA AAAAAAGTGC AGGGGGG

30

## (2) INFORMATION FOR SEQ ID NO: 150:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1569 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

60 GACCTGACG AGAGAGGCC TCTCTCTTCA GGGTGTGTGC TGTTTTCAG TGACGCTGCG  
120 GGATTAGCC AACTGGATC CGCGGCTGT GAGTCTGGA CAGTCAAGA AAGCGGTAGC  
180 CAGCGCTGTT CAGCAGGAG TAAATCTCT TGTGGCTTG GAGGCTCTC AGGTTCTGCG  
240 AGAGAGGCT CTTTCTGGG CTGGTGAGCC CTGTGACATC ATGAGACGA GTGATGAGAT  
300 GGATGCCAG GAGGAAGCA TCCATGAGAG AACTGTCTCC AAAAAAAGA AAGCAGAGAG  
360 ACACAAAGA GAATCGAGG GGGCTGGAGG AGAAGATAT CCAATGGATA TTTGGCTATT  
420 GCTGGCTCC TATATCCGTC CTGAGGACAT TGTGAATTTT TCCCTGATTT GTAGAGATCC  
480 CTGCACTGTC ACTTCCACTG CTGCTTTTGG GACCAAGTTG TACCGAAGCA CTACAGCTG  
540 GATGCTTCCC TGCCTTTGCG TCTGCGACA GATGTCAATG AGAAGCTGCG CTGTCTCGCG

60



600 GCTTGCTGTA TCCGATCTCT GTACCAATGAG TATGACCCAT TTGCTGCTGG AATCTCGAG  
660 AATCGAGCCA TTCCAGAGAG GACCCCGAG ACATTTAAGA ATTCCAAATG CTGACTTTTC  
720 TGTGTGAGAA AGATTGTGAG GAGACAGAG GAGACCAATG GGAATTTGAA CTTCAGATTC  
780 AAAAAAGT CCCGAGGT AAGAGAGAG TGTACAGAG GATTGAGAG TCCGCTGAG  
840 TACAGATG TTCAATGAA TCCAGAGAG GACTGCTGCC TACTGAGGT GACGAGCTTC  
900 AATTGACT TTATTCGAT TGTCAATGAA ATGAAATTTA CTCTGTTAC TATCAATGTG  
960 AAGAGAGCA TCCGAGATCA TCGATGAGA CTGATGTGCC AAGATTCCCC TGTCAATGAT  
1020 GGTGAGAAAC TCCGAGATGA ACGAGGTGTG CAGATCAATCC TGGACCGAGT GCAAGAGGTT  
1080 GCGCTTTTG ACTGTGTGCA TCTTCAGTAC CCAATTCCTCC TGAAGAGTTA GTTACTGCTT  
1140 CCGATCCCCT GGGGGCAGCC TCGATGTGAG TCCATTTAGTA ATCAATGCC AGTTGAGACA  
1200 GGGTGGCTGAG ATTGTAAATC TGTTTAGTAA TGTACAGTCT CTTCAGAGTTC TTAGGCTGCT  
1260 GTTACGGGAG GAGAAATGT TGAATCAAGA GGAAGAAACA CTACTATGAT TTATTAACAT  
1320 AATTAAATGT AAAAATTTGC ATTAAAGAG AGTGGCCGTG TTTTCTGTGT TAAAGCCCA  
1380 TTGTGTCTTA TTGAGTTGT TCTTAATCT TTATGCCAG TGAATTTGT TGAATCTGCT  
1440 GTAGGAGAAA ATTAACTCT TTGAATCTCC AAGCAGAGAA GTTTCAGAT TCCCTTAATG  
1500 ATCGAGAGAA CCTTACAGCC CTGAATTTGT TCGTTCCAGT TTAGCTGTCCC CTGAATTTCA  
1560 AGTGAATTTT TTCCCTCTTC CCTTACCCCT TCTCCAGAA TAAACAGAT GACGAGGTTT  
1569 GAAATCTT

40 (2) INFORMATION FOR SEQ ID NO: 151:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1540 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

60 CCGACGCGTC CGAGAGGATT GACAGATTAA CCAACATTT ACCCCGATG GCTGTGACC  
120 AGATTATGAC ATTGTGGCTC CAGATCAATG GCTGTGGCTT TATTTCCAGA TCGAATCTG  
180 TATTCATGTC CCGTGAATGAC GTCTGCTCTC GGAAGGTTTA CCAAGAAAC CCACTCTTAC  
240 CTGTGAAGCC TGTCTTTAAA GAGAGAGAA CTGAATGAA ACAGTGAAT TTACAGAAAG  
300 AATATGAGCC AAAACCCCTG GAGGAGATC ATCTAATGG TGTGAAGAC TCTAAGCTCC  
360 ATGAGCTTGA ACATGAGCAA GAGCTAATTT GTGCTGCCA GATGGTGG CCGTTCCGTA

420 CATTCCAGAA TGAATGGTTC TGTACTACA ACGAGCTGT GTTCTGGCT GCGATGGATC  
480 TTGCTTTGCT TTATTTGACT GTCTGGGACT TTGATGCTAT GACGAGAG TACGCTTACA  
540 CTGAGGACT GATGGGTTTC CAGTCAAT ATTGTATGAG GAGGAGAGC TATTAATGGA  
600 ATTAATGAAA CTGATGCTT TACTGGGTA CCGTGAAGAT GTGATTTGAT TCGGAGAGTTC  
660 TGAATCCAG AGTGGAGAG GTTCTGCTGT TGAATCTGTG TGTATCTCT GTATTCATGC  
720 CTGAGAGCCC CCGTCACTTG TCCGTTTCTC GTTTGAGAA TATCCGATCA AGGTTGATTC  
780 AAGAGAGTTC AATTACAGCT ACGAGATAC CTGAATTTAC AACGAGAA TACATGTCTA  
840 ATGGGCTTAA TTCTGCTAAT ATTGTCCCG AGACAGATCC TGAATCTGTG CCGATTAATCT  
900 CTGTGAGTCT GCTGTTTCCA GAGGCAATG CTGCTAGAAT CAGTCTTGG TCGCTTGAAT  
960 TACTGTGAC ACAGTTGCTG CAGAGAAAG TATTTGATTC TGAAGAGCC ATTATTAATG  
1020 GTTACAGAAA CTCCATGAC TATCTCTTGT ATCTCTTCCA TTTCATCAG GTCAATCTGAG  
1080 CTCAAAATCC TGAAGCTTTT GAGTGTCTG TATTTGATTC AGTCTCCCTT GTGCGAATG  
1140 GCGACATTAI GTATTTCCA TTGTCCGAAA ATACTGTGG AAGCAGATC TTGTCTGAG  
1200 GTCCGATGC AAAAGAGTTC AAGAGAGAA ATCAAGAAA TACATCTGTT GTTTCAGACA  
1260 GTTTACTGCT TGCATCTCTG TTACTAGAT ATAAAGACA CAGTCTTAA TTTTGTACTG  
1320 GAAATTTCA ATAAATGCT GAGTGTGTTG CTCTGTTTTT ACAGAGATG TGCCTTGAAG  
1380 ACTAAAGCT GTTACAGAA CCTAATGTC CAGAAATTTA CTGATTAAT TTCCCTTAAG  
1440 TTGAGGAGCA TGGAGAAAA ATGGAGAAA GGAAGAACTC AGTTTAAT ACAGAGACT  
1500 ATAAATGATA ACAGTGAAT CCCCTAATTC TGAATGATG ATCAATCTT ACCTAAGAA  
1540 GTGGTTAGTC ACATGAATTC AGTTAATAT TGAAGATTC

45 (2) INFORMATION FOR SEQ ID NO: 152:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1719 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

60 TACTATGAG GTCAATGGA AATAGAGCA CCAATTTACT GAGTCTAGAC TTTCAAATAT  
120 TACAGTTGAC ATGGTATGAC TTGTGTTGAG AACCTGAT GATGTGGACT GCTGGGCGAG  
180 TACAGAGCAT GTCTAGATTC ACCTTCTGTC CTGTAGATGC ACTTTTTC CAAATGCTG

5 ATCTGATCA ACAGGCTGTC GTTCAAGGA TGAATACAGG AATTCAGAGA TTATCCAAATG 240  
GTCTGGAGACC GGGCCCTCTAT GGAATTCATTT TCTACATATTT CCATCTGAAA CTTTAAAGAAC 300  
TCCCAATAC AGGACAGAC TTGGGAACAA ACACAGCCCC TCAGACACAC TTTCAGACAGA 360  
ATTCCATCAT CCTTGGCCCT CCTTCTCTAT TTGGAGCCCTG TTCAGTACTG CTGGCTCTGC 420  
TTCTTGCCCT GTTTATTCGG GAACATACCA ATTTAAGCTT AAGGTCCAGC AGTTGAGAAA 480  
AGCACTGTGG CAGTCAACAGC CATCTCATTA ATACACAGCC GCGAGGAGAG GCGAAAGAAC 540  
CTTTACTCCA GAGACAAAT GTGTAGACAC TGAATTCAGC AAGATTTTTC TATCAGACCC 600  
CAGGCTCTTAG TTTTCACTTC TAACTTCTGGA TGTACATTCG ATTTCATCC ACAGTACTAT 660  
TTAAGATTTG CTTAAGAAAT GTATCTGCAT GAATCTCCCTG GGAATTAAG GAATGGGAAA 720  
CTTAGACCA GACAGTTTTC CAAGATGTTT ACAAATTTCT TTGAAGAAC TTTTGTATTAT 780  
TAGACCAAT TTCTTCCAC TAACTATTT GTTTTATTTT ACATCTCTTA ATTAAAAACT 840  
ATATATGAA CTCTTAGAT ATTAGCAAT GTCTCTGCTA CCATTTCCCT AAGGTGTTCA 900  
GCTTTAACT TATCTGACT CAGTGAACCA CAGTAGGTAG TATGTTGTG GACCTATTTG 960  
TTTATCAAT GTAAATTTT GAGTCAAGT TTAATATTTT AAATCTCTGG GTCAAAATAT 1020  
TCAAGCCCTT AATGCAAGT CACTAAACCA AAGAAATGTT AATGAATTTG TTTCATTTTA 1080  
AAAAAAAA CTCTTAGAA AACTGTACTA AATCTGAATC ATGTTTTGAG CTGTGTTGCA 1140  
GTACTTTTAA ACATTTTCA CTACTTTTT TGAAGTGAGA AAGTATCAGC CATTTAGCAT 1200  
TTAAGTGGG GTATTGAGG CCGTAACTT AATGCTGGC TCAAAATTTT TCCCACTTA 1260  
CTTCTTATAC CACTATTTCT TTAAATGTTG CATTAATCAT ACGACTTCAA CACTTGAATA 1320  
CATATCTTAA AAATTTATTA GTAAAGCTGG TAGCCCTTGA AATGTCAGTG TGAATCTTAT 1380  
TATGTAGATA AATATATATA GTGGCTTTTC AGCACTGTCA CAGTAACTAT TTAATTTACAG 1440  
AGCTAATGTT TGTCTTAAAT TTTCAGACC CTAGAGGAGA GCTTTATACA ATTACCAATG 1500  
TGAAATTTCT TAAAGTGTAT ATTTTGTGT CCAATTTAT TATTTAAAA AATGTTACTT 1560  
TGTAAATTT GTATATTAAG AACTGTATAG TTACACTGT TTTCACTTG TGTGTGTTTA 1620  
TTGCTTAATG CTTTATTAAC TTGGAACACT CACTATGTTT AATTAAGTTC TTAAAGAAA 1680  
TGTAAATATT TTGTAAATAA AGTTAAATAT TTTAATGAT 1719

(2) INFORMATION FOR SEQ ID NO: 153:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 863 base pairs  
(B) TYPE: nucleic acid(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

5 GGACAGAGG AAGCCGGAC GATGTGCGCA TCACAAACCA CTTTGGAGTT TGGAGTGCT 60  
TCCCTTAGAG CAAAGGAAC AGCTCTCATT CAAGGAACCT AAGAGCTCTT CCTTAACTGG 120  
TAGGAGACA GCGAGAGCG GTTTCTGGG AACTGTGGGA TGTCCCTTG GGGGCGGAG 180  
AAACAGAG GAGATGCT CAGACAGTA ACTACAGCT GTGCTCTCT CTGAGATTCC 240  
TCTGCTGTC CTATGACCTC TTTGTCAATT CCTTCTGAGA ACTGCTGCA AAGACTCTTG 300  
TCATCAGCT TGTGCTCTTC ATCATCAGG ATATGCAAT COTCTTCAAC ATCATCAACA 360  
TTTTCTCAT GTTCTTCAC ACCTTGTCT TCCAGGCTGG COTGCTCAAC CTCTATATCC 420  
ATAAGTTCAA AGGAGCCATC ATCTGACAG CTGTGTACTT TGCCCTCAGC ATCTCCCTTC 480  
ATGCTCGGT CATGAACTTA GCTTGAANA ACTCCACAG CTTCATATGG ACAGATGAC 540  
TTCAATGCT GTTTGTATTC CAGAGACTAG CAGCAGTCTT GTACTCTAC TTCTATTAAC 600  
GACAGCCCT AAGACTAGG GATCTCACT TCTACAGGA CTCTTTGTGG CTGCGCAGG 660  
AGTTCATCA AGTTCGAAGG TGACTCTTGG TCACACTCAT GATFACCTTTT CTTTCTCTGA 720  
TAGAGGCCA CATTTGCTGC TTTTCAGGGG AGAGTTGGGC CCAATGCAAG GCGCAAAACA 780  
GTGCGATTT TCAAGGAAA GCGTTCAAAA TTAGGCTGT TGTTCAGCC ATTTCCNAGG 840  
AAGGGAAGG GTTTCCCTTTC CCT 863

(2) INFORMATION FOR SEQ ID NO: 154:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1101 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

5 AACAGAAA AAGATGATTT TCTTCTGAAA TTGTGGAACA TGAGGATTTCA AGTTTTTATT 60  
TTGTACTTAG GTCTGGAGG AACTCCCGAG TTCAACAAAG CCCCCTCTCT TCTCTGGAG 120  
CCAGGCTTG CCGTGGATC AAGTCCAACT GAAACATCAG AACAAATAG AGCAAAATAA 180  
GAATAGATG AATGACCCA AATATGCTT TTCTTGGGCG AGCATGTCTT GGAATGAGAA 240  
AGGTGACATG ACACAGGAG AGCAGATGG CACCCACAC AGAATACAGT GTGTGTATT 300  
ACGAGAGCC AGCAGTTGAG CTTAGGTCC TTCTACTTAC CTGCTATTGG CATTTAGGT 360

CGGAAACCTT CTACTGCCCC ATAGGCGAGG AAAAGTGGAA AGAGAGGAG GTCCTCTTAA 420  
GAACTGGGAG CAGAGCTTGA GGGCTTATGT ATGTAGCTGA GTGAGAGAG TACATGATGC 480  
TGTCTGCTTT GAAAGAGACT TTCTCTGCTT AGCTAGAGTA CTCCTTCCCTT AGTTGAGAGA 540  
ACAGTAGAG CAGAGCTCTG CTGAGTACCT CTGTGATGAC AAGGCTTGG TTAACTAGAG 600  
GTGATCTGCA GGTGTGAGAG TTATATGTGC CCGAGGCGAA ACAGAAATAT TAAATTAATA 660  
ATCCAACTTT AATAGATATC ATTAAAGAGA AAAAAAACAA AAGGCTCTGA AGTTGAGAGC 720  
CAGGCTTGCT GAGTATATGA GGTGATATTT CCGAAAGGGA ATCCAGAGCA AGTCCCTGCC 780  
TGTATTTTGT TCTTGAGAGG GGTGATCTTA GAACTAGAT CTAATGAGGA TGAAGAGGAG 840  
CAAGCCAGAG GTTGTCTGGA TGAGAGCGAA CGATTTTTAA GAAAAAGGA AGAGTTCTTT 900  
AGATAGATGA TTGTATATGA AGATAGCTAG TGAATACGAC TGACAGAGT CTATGATATC 960  
ACTATGTGTC GTTTTATGAA TAAAGATATC ATATCATCAT TCCCTTGGGG AAAATTTGTA 1020  
TTGAGGTATA AAAACAGAG ATTATATATA AAAATTTAAA GAACTCTTAA AAAAAAAG 1080  
CTGTGAGCGA ATTCCTCTGA G 1101

30 (2) INFORMATION FOR SEQ ID NO: 155:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2011 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1.1) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

40 CAATTAGACC GTTGAAGGCC TAGGTGTATT GCGAAGCCCC NAGGCTTAAG TTATTATTTG 60  
GCGAGGCGAA GGGCTTGAAG GCGAGAGGTA GCGGAGAGGT AGCAGGTGAG 120  
TTCTAGAGGC TGAAGAGTTT ACAGAGAGCC TGTGTCAATG CCGTGTATAT AAGAGAAAC 180  
CAGGCTGTC CTGAGTACCT ACCAAGCTTG GTTTGACAAA AAGCAAGGTG GAGAGCTATT 240  
TTGTAGCAGT AGATAGATCA CACTTACCTG TGGGCGAGTA TTGTGAGTGT AGTGTGAGTT 300  
GTTTACAGTG ATGCTTCCCG TGGCGACGAC AAATTGTGTA CATAGCTTTC AGAATGATATC 360  
CAAGCTTTTC CCGAGCTCCC AAGCAGAGCC TGGTCTTGAAG CTGTGTATAT ATGTCAATAT 420  
TACGTTTTTT AATATATGAC TTATGATTTCT GTTGTATTGA TTTTAGAGCA GTGTATGAC 480  
CTTCAATTTA ATAGATGTGT GTGATAGAG ATAGACATAT ATGTGTGTGC GTATGCAAT 540  
ATCTCTATTC TGTAGTTTCC AAGAGTTCAG CTGAGAGGTA TGAAGTCTTC GAGCCGAGGA 600  
GAGAGCTGTC ATCCCTCTCA ATATGTGTTG CAGAGAGTAT TAGTATGTCT TCTTATTTAA 660

TATTTTCAAT TCGAGAGACT GAGGAGAGCC TGAATGTGTT TGCTGTGCTT TGTGAGAGTA 720  
AGTGAAGGTC TTGAGATGAC TTGCTGTGTT CTTGAGAGTG CACTTGGAG CCAATCTGTC 780  
AGATATGAC CCGCTTTTGG CTGTGTGATTA CTCGTCTGCT GCGTGTGTGT GTGTGTGATTA 840  
GTGATCTTGG CAGGCTTCC ATGTCTGTGT TTGTGATTTT TGGGATGAGT GGTGAGAGCA 900  
ACAAATTCGA GTTGTGTTGA GGGCTGTTGG CCAATGATAG ATCACTCTTG TTGAGCTGCT 960  
ATGTCTGCTT GCTTCTGCTT TTCTCTGCTT TCTCTGGA AGAGAGAGAG AGCTGTGCTCA 1020  
GGCCGAGGCT GATGTGATAG AGCTGAGACT GGTGATGACC GTTCTTGAAG CAGAGAGAGG 1080  
AGATGTCAAT TTATCTAATG TCCATTAACA ACAAATTTGC AAGCAAGACT CCTGTGCAAT 1140  
CGAGAACTG AAGCAAGATA GGGTTATGAC TTGTGAGAGA ATATGATGCT GCTTGAAGGAGT 1200  
AAATGAGGC ATGAGATTTT CTGTGTGAGAG GCGAGAGTAA GCGTGTATAT TGTGAGAGCT 1260  
TTTTTTTTTA GCAATTAATT TTGAGCTTGT TCCCTGTGAA GTGAGAGAG GGTAGAGCTT 1320  
TCCCAATGT AAGCTAGAT CTGAGAGAG CAGAGCGAGT TATATGTTCTG CTTTCCGAGA 1380  
GAGAGAGAC TTTTGAGAAC CAAATGATCT TAAATGTATAT TGCCGAGCCC TGGCTTTTCC 1440  
GGTTAGAAA TTGAGAGTAG GAATGATGTT TGAAGAGAG TGCTTGAAC CAGTGGGTAA 1500  
GAGAGAGAGC TACCTAATCT CAGATATCTG CAGCAGAGCC AAGCAGAG CATTACTTAG 1560  
CAGAGAGAAA ATGATGTGAT TTGATGTGCT GTGTGTCCAA AATGAGAGCA CGATGTTCTT 1620  
TGAAATCTG CCACTCTGAG GCTGAGGAGG GTGTGTGACA CCGTGTATAT CCGAGACTTT 1680  
GGAGAGGCTTA GCGGAGGAGG TGACGAGAGT CCGGAGATTT GAGAGAGCC TGAAGAGCA 1740  
TGGAGAGAAC CCGATCTCTA CCAATTAATA CAAATTTAGC GGGGCTGTGT GGCATGCGGCC 1800  
TATATCTCA GCTACTGAG AGGTTGAGAG CAGGAGATAT GCTTGAAGCC TGAGAGGCGG 1860  
AGATTTCCAG TTGATGAGAG GATGTGTGCA TTGCACTTCC GAGCTTTGGG GCAAGAGAG 1920  
CAAAATTTCC GTTCTAATTA ATGTGCGAAT TGAATATCA GCTTATGAT ATCTGTGAGC 1980  
TGAAGGAGAG GCGCGATGAC CAATTCGCCC TATGAGATATC GTATGCAAT C 2011

50 (2) INFORMATION FOR SEQ ID NO: 156:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1981 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1.1) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

60

5 CCGTCAAGCTT GAGCCCTTCA CCGCTCCGAG TTCCGCCGAG GTTGGCTCC TTGCAATCCT 60  
TTTCTTGCTA TCAAGCTTTG ATTGAGGAA CACCCGCTC TTGTGCAAC TCAATATGA 120  
GCTCACTGT GAGGAGCAG TCGGCACAG CTCACGCTTA TGTCTATGT TACCCGCCAA 180  
GACCGCAAA ACTCTGTGT GACACAGAT GAGCCCTCAG CCAAGATCT GAGCAGCTT 240  
GTGCTCTCG CAGCTGAGC CTGCCCCATG TTAGAGAAC AGCTCATGA TCCCGGGGA 300  
CGTGGGACA TCAGGACAT GTTCCGGCCG CCGTTGGACA TTACAGACT GGTGATTCG 360  
CTGTATCTC CCATATCC GCGGACGCG AGCTTGTG ACTGCGAGY TCGCTCTTC 420  
TGCCTGGGCT TCGTAGCCA GCGGGGGCTC TCATCCCTGA TCCCGGTGCT GGTATATG 480  
CGTCTCAGC TCTATCTGAC GCACTCAGG GAGGCTTTTG GCGATTCGC CTTTCTTC 540  
TATGACAGC ATGCTGAGCA GTGTATGCT GTCTCTGGA AGCCGACAG CTTCAGCGG 600  
CAGCCCTCA AGGCTTCAG CACAAAGGG GCAATGTGA TGTCTCAGG TGGGAGCTA 660  
GTATGCTGC CCAATGTGA AGCAATCTG GAGCACTTTC CTGTCTGCG TGAAGGCTG 720  
GTGACAGT TCGAGCGCG AAGTGAGGG TCGACTGTGT GATTCGAGT CTGAGCAG 780  
CTGTAGAGG AGAGCAGAG ATTGAGCTC TAGAGCAAGA TGTGAGTAG ATGAGCTCCA 840  
CGCTCTTGG AGATGATTC TCAATGAGG GCTGTCTGCG TGAATGCT GATCATCTC 900  
CAACAAACC CAGCCCTCAAC TTCTCTCTG ATGCTCCAGC ATTGGGCGAG GGGCATGCTG 960  
GCCATGTAG TCTCTCTGCG CTCACATCC CAGAGAGGA GTGGAGCCA GCTCAGAAA 1020  
GGAATGAGC CGAGAGATC CATCACTA TTAGCCCTGG GCTGTGACT CCGTGGAT 1080  
TCCACTCTT TTCTTAGCT TCTTCCAGA ACAGAGAGG GGAATGTGCT CTGGAGAG 1140  
CTCTCTCTC TTCTCTGCT CAGGACTGT GCTAGACTT AGCATGCTT TCACTGCA 1200  
GTGAGGCTT TAGATGGAG CAGCGGAAA TGTGGGCTT CTGATCACA TCAAGGAC 1260  
TGAGCATGG AAGGGGCTA TATGTGTAG ATTAGACAC ATTGAGGAG CAGATGCC 1320  
TCTCTGTG ATGCCACTTC CAGAGGTGGA GACAGTGGAA AAGAACCGAG GACAGGAG 1380  
GATGGGTAG GTAGAGGGGT CAGGGGACTG GTACTACCC AATCTGGAG AGTGTCAAAA 1440  
AGCACTGGG GCTACCGTT AGCTGCTCT GCGCTGGCTG TTGTGCGCTT CATGTCAAAA 1500  
ACTGCCACTA CTATGTACT GCAATGGGCT TCGAGAGAT GGGGAGACT AGTCTTACT 1560  
CCCGAGAGC TCCAGAGGCT CAGAGAGAG AATGCTGCT CTTTCTAGT TGTCTTAC 1620  
CCACTTTCTG GTAGGCTCT TCTTCTCTT AATCTGCTT GTTTTCCAG ACTGAGCTCA 1680  
AATAGTCCG CTCTTAAAG CCAATCCCTG CCGCAGCTT GAGGTGATCT TTCTCTCTC 1740  
TGACTATTA GAGGAGTAC TGTCTGTCA GTTGTCTTGG CAGGAGACA CAGTGGCTA 1800

5 AATCTATTG TTTGAACTC TCAATTAAA TTAATGCA GCTGGCGTG GTGGCTCATG 1860  
CTTGTATCC CAACACTAG GAGTAAAGG GAATCACTTG ASCYAGAGAG TYCTAGACCA 1920  
5 ATCTGGGCAA MAGAGAGCC CCACTCTTT TAAATAAAA GTTAATTCG TTAATAAAAA 1980  
A 1981  
10  
15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 915 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(2) INFORMATION FOR SEQ ID NO: 157:  
20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:  
GAATTCGCA CAGGCGCGC CATGGGCTC GTCTTTGCG TCTGGGTGT ACTGCTGGC 60  
GGCTTCTCG CCGTCTGCG GTTGGCCAAG CTCGCGAGG AGATCTCGC TCGAGTTTG 120  
GAGCGATGA ATCGCTGTT CCGTCACTTT GCTGAGTGT TCCGCTGAA GGTATTTGC 180  
TACCGCCAG ATCCCTGAA CTACCAATA GCTGTGGCT TTCTGGAAT GCTGGCTGG 240  
TTCTCTCTG TCATGGGCGC ACGATGCTG CAAGAGATCA GTAACTTGT TTGATTTCTG 300  
CTCATGCG GGGCTATTT CACTTGGCA GCTCTGAAG AGTCACTAG CAGCTGTATC 360  
CGAGCATG TCTGCTGGG GTTCTGCTG CTGCTGANTG TCGGCGAGT CTTAGCCGAG 420  
ACTAGAGG TGTGAGACC CACTAGGAG AAGACTCTAA GTACATTCAA GGAATCTGG 480  
AAGTAGCCA TCTCTCTC TTATGCCAT GCAGCTGTCA CAGAGGAGC ATGCTAGAAC 540  
ACAGATCTA TCATCTGTT ACCAGTATA TATCGAGGT CAGCCATGT TGAAGAGAC 600  
ATTGTCTA CCGCGACTG CTTCTCTTT TTAGCTTAC TACTCTTTG TGAGGATTC 660  
ATGTATGCA TATTAACAT CCGTATGTA TATCAATAA CAAATTAAG AGAAGAGAA 720  
TTTAATCAA CCAAAATCT GATCCGCAA ATACCACTT TTAATGCTT GGTGTAGTA 780  
TACTCTGAA CTTTTCTG TCGCTTAAA CAGATATA TTTTTTTAA ATGAATATA 840  
50 AACCATAT CCAATTTAT TTCTCTCTT TAAAGCTTA TAACTATAA MAAAAAAA 900  
AAAAAAAA CTGCA 915

(2) INFORMATION FOR SEQ ID NO: 158:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2117 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

AGAGGAGAG GAGGTGGAG GAGTCCGAG CATGAGACTG GAGCGGAGAG TGGTGGAGCT 60  
GCTGCTCTG GCGCGTCCG TGTGGAGAG GGTGGAGCC ATGAGAGCTG GAGTGGAGCT 120  
GAGCGGAGCT CTGAGCGAGT ACATGAGCC GGTCTCTGAC TGGCTCTG GCGAGTCTG 180  
GCGGAGAGAG GCGAGAGCTG GCGGAGAGG AGTGGAGAGAG GAGTGGAGAG AGAGAGCTTT 240  
TGGAGAGAGT CTGGAGAGG AATGAGCTT AATGAGCGTG TTGAGTTGA TTAAGAGAGC 300  
AAAGCCGAG AAAGCTCTG GGTCTCCCT GAGAGGAGTG AGAGGAGAG GCGAAAGATT 360  
GTCAGAGAG ATGATGGAG AGAATATTGA GAGAGGAGTG CTAGAGAGTG ACTATGTCGA 420  
CGTGTGTG GCGAGATTC AGTTTCGAG TGGTTGAGC ATGAGCTCTT ACAGAGATGA 480  
GTGAGAGTG TGAATGGAG GAGAGTGGAG TGGCTGTGAG AGGTGAGAT TCAATTTTGA 540  
TGAATGAGAT AAGATGGAG GAGGCTCAT AGATGAGATC AAGGCTTCC TCGAGTATTGA 600  
TGAAGTGTG GATGGAGCTT CTGACAGAG ACCGAGTCTT ATATTTCTGA GCAATGCTG 660  
AGCAGAGAG ATGAGAGTG TGGCTTTGAG TTTCCTGAGAG AGTGGAGAG AGAGAGAGAG 720  
CATGAGCTCT AAAGAGATG AAGAGAGCTT GTCTGTGTG GTTTTCAGATA AGAGAGAGAG 780  
TGGCTCTG GAGAGAGCT TAAATGAGAG GAGAGATC GATTAATTTG TCCCTCTCT 840  
CCCTCTGAG TACAGAGAG TAAAGATG TGATCGAGTG GAAATGAGT CCGAGAGCTA 900  
TGAATTTGAT GAGAGATG TACAGAGAT GGTGGAGAG AGAGAGATTT TCCCGAGAG 960  
GAGAGAGTT TTCTCAGATA AAGGCTGAG AAGGTGTGTC ACCAGTTGAG ATTATTTACTA 1020  
CGATGATGAG CATGAGATG TGGAGAGAG AGTCACTGCC TGGAGTTGAG AAAGAGAGAG 1080  
CATGAGCTG TGCAGAGCT CCGAGCGAG GTCTTTTCC TGGAGAGAG ATTCAGTGA 1140  
TGTTCCTGTT TGAATGAG GAGATTTCT CTGAGATGTT TTGAGAGCC TGGTGGCTG 1200  
AGGAGAGAG GAGAGAGAG GAGAGAGAT GAGAGCTGCC GAGAGAGAG AGAGAGAGAG 1260  
AGGAGAGCT GAGAGAGCT ATGCGAGAG TCAATGATTT TTAAGAGATA TGTTTAATTT 1320  
CGAGTGTGTT CTGTTCAGAG GAGAGAGAG TAAAGTTTAT TGAAGATG TGAGAGTTAT 1380  
TTAAAGATG TTAAAGATG ATGAGAGAT GGTGAGATTC TAAAGTTGAG GCTGTGTG 1440  
TGTGTGTGTT TTAAAGTCT CATGATGAT ACATGAGAG TGAATGATCT TTAAGAGAG 1500  
TGAAGCGAG GAGAGAGAG TGGAGTTGTT TCGAGAGAG GCGAGTCTG GCGAGTGTG 1560  
TGAAGAGAG GCGAGAGCTG TCGAGAGAG GTCCGAGAG AGTGGAGAG GCGAGTGA 1620

GTGTTAGAT GAGAGAGAG TGAATGAGAG GAGGCTCTGA GCGAGAGAG ATGAGAGCTGA 1680  
AAGAGAGCTT CGAGAGTTT TCACTTGAAC AAGATGAG GGTGGAGAG TGAATTTGAG 1740  
GAGAGTGA ACATTTGTA GAGAGAGCT TTGAGAGAG AAGATTTCC AGTGTGAG 1800  
TTCAGAGTGA CTGCTTTG GAGAGAGTG GCTTCACTG TCGATTCGA ATGAGAGAG 1860  
CGTGTCTG GCTGAGTCT AATTTATGAG AATTTGCA TATTAATCT TCTTCATTTA 1920  
TACTTTATTT TGTAGAGTG ATGAGAGAT CCGAGAGCT CTGTAAGAG TTAGCTGCC 1980  
CTTTCAGAG TGTGAGAG GGTGTCTT TTATTAATG GAAATGCTA CCGTTTGA 2040  
ATTAATTTT GATGTGAG AAAAAAAAA AAAAAAAAA AATATCCG GAGGAGGAG 2100  
GTAGAGATTT TGAAGCC 2117

(2) INFORMATION FOR SEQ ID NO: 159:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2195 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

TGTTCCTTGA TCCCTTTCT AAAAAAAAA GAAATCCG AGGATTTTA GCGATGCTG 60  
TGTGTGAG TGTGTTTAT TGCAGAGTA AATGAGAT ATAGGCTTT CATTTCCG 120  
CAAGGCTTT ATTGGAG TTAAGAGAA TGTGTTTCC AAGAGTTGAG TTAATTTCT 180  
CTCTTCTTT CTCTTCTTC CTCCCTTTT CCGTCTGAC CCGAGAGCTT ATTGCGAAA 240  
CATGAGTGA GAGAGAGCT TGTTCCTGA CCGTGAATA TGAAGAGTG CTAAATTTGA 300  
GAGAGAGCT AGTTTGGAG TTAATGCTG GATTTTCCCT ATGCAATGAT ATTAGAGAG 360  
AAAAAAAA CTGTTCTG TGTACTTGA GTCTTGAAG AATGTGCCC ATAGTTTAT 420  
GAGAGATTC CAAGGCTTT AGTACAGCT GTATTCGAA ATGGGAGAGC GAACTCCG 480  
GAGAGAGAG GCTGTGAG GAGTACGCT TGAAGTGA AGGAGAGAG AATCTGAG 540  
TGAAGTTT GAGTTCAG TGTATTTGCT CTGTTATTC AAGAGAGAG AGAGAGAG 600  
AGAGAGTTT GAGAGAGTG AGGTGAGC AGTGTGAG AATATTTAG AGAGAGAT 660  
CTGAGAGAT GTGAGATTC CGAGATTC CTGAGAGAT TCCCTGAGC TTTTACAGAG 720  
GATGTGCTA GAGTTCAG CTCTCGATA TGAGCTTAG AGAGAGAG TAAATGCTG 780  
TACTGAGAG ATGAGTTGA TTCTTGGCC TCGAGATG ATGAGAGAG TGTGTCTCT 840  
CTGTTTCT AGTGGAGAG AATGAGTGA GCTTTTGA GCTGTGAG GAAATTTGAG 900

5 TTTCTCATG GTGACTATCA AAACAGTTT CTGCATGAC TGAGCCGCA GACAGAGCT 960  
GGAATCGTTG TCACCAATCC TAACGACTCA GTGTTTCTCT TCATTGACAG ACAGCACTTG 1020  
CAGACTCCAA AAACAAAGC TACAACTTC AGATTATGCA GCATCTGCTT CTACTGCCA 1080  
CAGGACAGC TCACCCAGT GGGGAGTGG CACATAGAG GRTCACTCC GTCTTATAT 1140  
GCCAGAGTAG AGTACTGACC ACCAAATGG AGAGATCAG AGAATGCAGC AGCAGTTT 1200  
TTTCTGTTT TCTTACACT TTATCTTTC AGAGTTTAA GAAATGAC TCATGACAG 1260  
AGACTATGC ATTITGAAC TTGTTCACC TGGATTTTTT TAATCATTT TTATCTGAA 1320  
ACTTAACAA AATATAGAT TGTGTCAGG ACTGTGTGAA AGAGATGCT TTGCATATTT 1380  
GCTGCACTGC ATCAGTATCT TACTAATAAT GTGAATGAA AGCACTATG TACACTGAA 1440  
TGCCTAATG TATCTGAAG CACAGGTGA TACTCATTTT TATGTCCTC CCAATTTGTC 1500  
TGTTTTTTC CTCCTTGACA TCTGTCATCA GTATTTAGAG GTGAGAGCT GAATGTACA 1560  
GGTATAATA ACATTTTAA AACATTAAC TTGCTATTA TCACAGTTCT TCCAGGAC 1620  
TCTCAGATAC ATTCATATCA CAGACATGG TTTAACAAA GAAATACAA CCATGGGAA 1680  
GAAATCTTAA ATGAAAGAG CATCTCATG TACGCAATTT TCCCTCATAT TTACTGCGC 1740  
CATGTTTGT TCCGTGACT CATGTATTTT TTTTTTCAG ATCTCTTTCC CCAGTTGCT 1800  
ATTGTAGAG TATTTCTGCT GGTGTGAGT CAGTTATACA CATTAAGCA GATCTGGAT 1860  
CTGAAGTAG TATTAAGAG CTATTAACA GAATATAGT CATAGCTGCA GAACCATGA 1920  
TAGGTAGAG ACTTTCTTT TGTTTTCTT TTTGTTTGT TTTGTTTGT TTTGTTTGT 1980  
CAGAGAGAG ATTTTTATTA CAAAGAAAA AATTCAGTG AATGTGACAG AATCTGCT 2040  
TTTTACACA TCTTAAGAA AACCTTACA AGGGTGTTTT GGAATAGAAA AAGGTTATA 2100  
AAGTTGGAT CTTAATTTGT AATATTAAC ATTAGCTGTC AAGTTCTTAA AGCAGAGCT 2160  
CATTTGTC ATGAACATA AGGAAGACT ACTGTATAGG TTTTTTTTTT TTCTCTTTT 2220  
AATGAGAAA AGCTTTGCT TTAGGTTTC ATACTTTAT TGGATTAAT CTGAATGATC 2280  
CTACTCTTT GGAATTAAC TATGCTTAC CAGTTTCCAA TTGTATTTAG CTTCGTGTTG 2340  
GAATTTGAAA AAAAAGAAA AAAGAAAA GAAACCTAA ATTAATATAG TGAAA 2395

55 (2) INFORMATION FOR SEQ ID NO: 160:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2120 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

5 CCCCAGATAC CCGCTGACT AGTGCCATC ACACCTCTCG GTCTCGCGC CCTCGAGGC 60  
TAATGAGAC GCTTGGCAA AGCAGTAAAC GAATTTCCGG GTGACCTTC GCTTTAGGC 120  
TCGTAGTTC TTCCGCCAA CCGAGAGAA CCGGAGAGC AGTTTACAC ACCGCCGTC 180  
GTGTTAGGG CGGCGCCGC TCGCGCGCA TGTTCCTCT TTTCTGTTT TCTCAGAGT 240  
GCTCTGCTA AGCGGTCC CCGACGCAC CATCTGTTTC CATCCCGCC GCGCAGGCA 300  
TTGCGATTT TCGAAGTGG CAAGTTCTAT GACACCCGTG ATCAGAGACA ACCCTCAGG 360  
CTGGGTCCC TGTGGTTTC CCGACAGTT TCGGATATG CCTTACCAGC GTTTCAGCA 420  
AGGAGTCGG CTAGGAAGG TTGCAAGCTG GACAGAGGC ACATACCAG ATATAGAGTA 480  
CAATATATG TACTCTCTC AGTTGCTGG TCGAATCAA TATGCTTAT TCCATGAGGA 540  
GGATGAAGT AGCTTCCAG TGTGTGATAC AGCCCGACA CAGAGAGCG CTTACACGG 600  
GAATCGAATG AGATTTGCCC AGAGGAACCT CCGCAGAGC AAGATGCTC GGAACATGT 660  
CGATTTCAAC CTGCAATCC TGCCTAAGAG TCCCAACAG AAGAGAGAG AACCATTCG 720  
ACTCGAGAA AGTTTCCAG ACAAATTTGG GTTATGGAG AATGGGATC AGAATACGA 780  
GAACCCCGA GACTCTTCAG TTGAATTTGG TACTGATGG GAATGTAAG AGGAATGGA 840  
TTTTCTCAG TTGATGAGA TCGCTACTT GGAATATCA GAGCCACAGG ACATGAGTG 900  
TTGTGGGCG CTAGATACT AGCAGAAAGC CTTTGACCG ATCACAGA GAGTGAAGA 960  
GCCATGCGG ASATNCAGC GCATCTTCCA CACTGTACC ACCACAGAG ACCCTGTAT 1020  
CCGCAAGCTG GCAAAATCT AGGGGAATGT GTTTGCACT CATGCCATCC TGGCCAGCT 1080  
GATGAGTGT ACCCGCTCAG TGTATTCCTG GGAATTTGTC GTTCAGAGAG TTGGTTCGA 1140  
ACTCTTCTTT GACAGAGAG ACACCTCCA CTTTGACCTC CTGACAGTGA GTGAGACTGC 1200  
CAATGAGCC CCTCAGATG AAGGTAAATC CTTCAATTCA CCGCCGAACC TGGCCATGA 1260  
GGCACTTAC ATCAACACA ATTTCTCCA GAGTGTCTTG AGAATGGGGA AGGAAGATA 1320  
CACTTTCCC AACCAACC GTTTTGTGGA GACAGATG GATATAGATG AATTCGCTC 1380  
TGTTCGTAC CTTTACCGCA GTGGAAAGCT TCGAGATGAT ATTGACCTTA TTGTCCGTTG 1440  
TGACAGAT GCGTCAATGA CTGAGGCCAA CCGGAGAGTG TCTTTCATCA ACATCAGAC 1500  
ACTCAATGAG TGGATTTCCA GGCATCTTAA TGGCCTTAC TGGCTTACA AGCTGAGCTC 1560  
TCAGGAGGG GCTGTCAATG CACGAGAGT GAAGACAAAC AGCTACAGT TGGCCGCGTG 1620  
GACCTCTGT GCTTTGCTGG CTGATCTGA GTAGCTCAG CTTGTTATG TGTCTCGTA 1680

CGACGTGAAA GACTCTCCAC GGCAGCTCAT CCTAGGACCC CAGACATTCGA ACCCTAATGA 1740  
GTTTCGACAC GAGATCAACC TGACCTCTGGA GAAATCCCTGG GGCATTTCAC GCTGCTCATAT 1800  
TGACATCTGC ATGACCTGAG AGAGAGGCGAA ATACCTCATAC CTGAGAGACC CGAACAGACA 1860  
GCTATACCTG GTCTACAGACC TCCCTGATAG CACTCTACAC TCTGATGAG ATGAGAGAGA 1920  
ACAGAGAGAG GAGAGAGAGG AAGAGAGAGA GAGAGAACT TAAACCATG ATGTGAGACT 1980  
GAGATTTGTC GTTCACACGA GACTACAGAG GGCCTTGATG CTTAGTGAAA TGTGTGTCTA 2040  
ACTGCTCTGC TGACATTTAG CAGATGAAT AAAATATATA TCTGTTTACT CTTAACAAAA 2100  
AAAAAAAAAAAAAAAA 2120

## (2) INFORMATION FOR SEQ ID NO: 161:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (k1) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

GGAAAGTGAAT GTCTCTCCAG ACCAGAGACA ACCAGAGCAT TCTCTAATGA GCTGACACGA 60  
CCTTCACCTT CACTCTTAC TGAGACACAC AGAGACAAAA GTTCTCATCT AAACTGATATG 120  
CGAGAGATAG GGCCTTGTTC ATGAGACAGA ACTTCTTCCA GCGAGAGGAC AACCTCTGAC 180  
AATGACACAA GTGAGAGAG CTGTACTCGA CCCCAGTCTT GGCATTCCTT ACTCTCATAG 240  
GTTTCAGTGT TGACACAGAC AAAATACAGT TCTTGTGTGT ACCAGAGCTG GAGAGAGACC 300  
TTTCAGTGGC CTTGATGTC AACCCAAAGC ATGTGCTGTG CAGAGAGTTC TGTGCTCCAG 360  
GTGACCTGAC GGCCTCTGGA TGCCCTGAGG TTCTCTCATG AAAAAATAGA TGTTCATGGA 420  
AATGACAGAG CTGAAAAATAT GTTGTGTGAT CAGAGAGACC AGAGTCAAGT GACTTTTCCA 480  
GGCTAATGCT TGCGATTTCCG CTATTTCCCA ATGTCCAAAC ACCTGACCTA CTTGAGAGGC 540  
AGCAGAGACC CTGAGAGAGG GAGACTTGAG TTCTATTAGA TGACCTGGA CAGAGAGTGC 600  
GGGCCCCCCC GCGCGAGGGA CTTCCAGAGC CTGGCTTACT GAAATCTGAA GTGGCTTAC 660  
GGGTTTTCAC CATTGACAAA TTGCTCTCCC AAAATATGAG ACATCATGAA GCAAAAACAG 720  
AAATTTTGTG ATGAGCCGAG GGCCTTGTG TGACCTGAG GTGCATGAT CAGGCCCTCA 780  
GAGACCTGAC AGAGATACCT GAGAGTGTG ATGGCCCTCA GGTATAGAGA GAGGCCGCCC 840  
TTACGCAATGC TGAGAGACAA CTGAGAGCT TTGCTCCAGG ATGTGCTGT GTTCTCATAT 900

## (2) INFORMATION FOR SEQ ID NO: 162:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1003 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (k1) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

GGCAGAGAT GAGGAGACAC CAGTCTTCT AGGAGAGGCT GGGTGTGCT CCCCAGATGA 60  
TGACCTCTGC TTACTTACT CTCCGAGAT GTTAACTTT CTATTTCAG CCTGTGACAC 120  
CTGTCTAGAC AACCTGACTT CCCCATTGAC CCGTGTGGCT CGACAGAGAC GTGGTGGCCC 180  
CCAGAGGCGA CCGCTCTTT CTGATCCCTC TTTCCTTAC ATGACTTGG GCTTGAGTCT 240  
GCGAGAGAAC CTTCCTTTTA GCTTCACAC CAGAGAGAGA GTTGAATAG ACCTCCCCGC 300  
CCCCACACA AGGCTGAGAA CAGAGAGAT GTGTGAGAG CAGAGTCTT GTGGCCCTCT 360  
CCAGAGTGT TTGCATGAT CACTACTGTC TTCTCTTGT ACCTAATCA TCAATATCT 420  
TCCCTTCTCT GTGGGAGAG GAGAGCTGC TGGGTGAG CTGACCTGC CACTGAGTT 480  
GCGAGAGAG GATTAATGAT GAGCAGTCTT CTGTGAGAG CTCTGATCT AACCCAGCCC 540  
CTAGATTCGA GAGTGTGAT AACGTCCAT GAAACAGAC CTTGACAGCA AACCTGAGAA 600  
TGCTTGGAGG TGAGAGAGAA CTTGACTTCT TCTTTCCTCT TCCCTCTCC AACATTATG 660  
GAACTCATC CTGTTAGAT CTTCAGAGCT TGTTTTCCCTG CTGGAGGGA CAGAGAGAGA 720  
ACAGAGAGG AGGCTTAGA AGAGCAGCC CTTCCTTGTG CTCTGGGCTA AATGACTTTC 780  
ACCTAGAGTA AATGAGAGA CAAAAACCT CTGATTTTA ATTTCATTA AATGTTAGAA 840  
GTATATATAT ACATATATAT ATTCTTTAA ATTTTGAAT CTTAGATAG TCTAAAAATC 900  
CAATCCCTCT GGCCTGAGC CTGAGTGA CAAGAGAGA AAATGTGT TTCAATTAGG 960  
AATGTAATTA AATGATGAA ACTGAAAAA AAAAAAAAAA AAA 1003

## (2) INFORMATION FOR SEQ ID NO: 163:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2196 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (k1) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

AGAGAGGAC AGACAGATGT GAGGTTCAC AGAGAGTGG GAGAGATAC CAGGAGCTTC 60

GGGAAACATC AGCATATGCA TCACCGAGAT GACCTCTATG CTCAGCAGAT GGAACGAGAA 120  
ATGAGGCACA AACTGAAAC AGCCTTTAAA ATTTCATGTC AGAAGTAGA GGCCTCTACT 180  
AAGAGGAAAC TCGAATTTGA AGTGCCTTTT AGGGACTTGG GATTTAAAGG AGCTCCCTAT 240  
AGGAGTACT GCTCTCTTCA GCTCCACTAGT AGTGGCTGGT TAAATGCTAC GGAATGGCCA 300  
CCTTTTGGTG TCACATTTGA TCAGGTAGAG CTGATCCACT TTGACCGGCT CCAGTTTGCAC 360  
CTGAGAACT TTGATATGCT AATGCTCTAC AAGGACTTACA GCAAGAACT GACCATGATC 420  
AAGCCATTC CTGTAGCCTC TCTTGAACCC ATCAAGGAAT GATTGAATTC CTCGACCTG 480  
AATATCAGAG AAGGAGTACA GTTCTCTAAC TCGACTTAAA TCATCGAGAC CATTTGTGAT 540  
GACCTGAGG GCTTCTTCCA ACAGGTGGC TGTCTTTTCC TGGAGCTTCA GGGTCAGGG 600  
AGTATGCTG AAGAGGGGA TTGAGAGTCT GAATTTGAAG ATGAGACTTT TAACTCTTCA 660  
GAAGATGACT ATGAGAGGA AGAGAGGAC AGTATGTAAG ATTATCTATC AGAAGCGAA 720  
GAGTCAGACT ATTCTTAGGA GTCATTTGGT AGTGAAGAG AGAGTGGAAA GGAATTTGGAT 780  
GAATCTGAGG AAGAGGCTCG AAAGGCGGAC CGAGAAATTC GTTAGAGGA AGAAGAGAA 840  
CAAGTCTGAA GTATGAGCGG GAGAGGAGG GCATCTGTGTC ACGTTTGGG CCGTGGCTCT 900  
AACCCTGTTT CAGACACAG CTCCTGCACC CCGAGMAA AGAGGAGTA ACTTCTGAC 960  
TTTGGCCCTG AGCTCCATTC TTCTCTCAGC CAACCCCTCA AATTTTACA TCACATAGAA 1020  
ACTGTATTTT TCCTTTGGT TTCAATTGAA GTTTTGGCAT TTGTGTTTAT GGGTTTAGGG 1080  
GGCAATTTGT GTGACCAAT CTACTCGGG AATTCCAGGC CCACACAGAC AGTGTCCMAT 1140  
GGCCCATTC AGATGGCAG GAGAGAGGTG TTCTTGAGA CAGGAGAGG CTCGCCCTGT 1200  
TAAATAAAT TGTTTTCAATC TTCTCTCTTC CTGTCACTTT CTGCGAGAC ATTGATGGCT 1260  
TCGTGACATC TATTTGTTGT CTCAAAGCTG TATTTCCAG ACGTGTGAC AGGTGACCC 1320  
TTAATTACCC GTATCATGCT TCTTGAACCG CACATTCMAT CCTCCAACT ACCCTACTGC 1380  
CATGACCTTC GGCACATCTC TAAGTTTATC CTTTGCATA CTCAGGTTTC TCGGAATTT 1440  
GCTAATGGTT GTGTAATACC ATACAGCTTG AGCCAGTGAG GCAATTTGGG CTGTGCTCTT 1500  
GCTCTGAGTT TTCTGCTCTT CTTGCTCTGT GCAGATCTG AGTATATCT GTTGCCTTGG 1560  
AAGACATAG AAGCAGTAT ACTGCTGGC TGGGTTATTT TTCTCATACA ATGCAACMAT 1620  
GGTACATGA TAGAGGCGAA AATTCOACT GTCTCTCTTT TTTTCTATA TATCTTAGGA 1680  
AGATATATCA GTTTGTGCT CATGTACCG TTCTAGTGAA ATGTAGAGGA AGGCTCAAG 1740  
GAGTCACAT TTAGATCTG AAGGACAG TCATGCTTG GGCCTAGMAT ACCCTGATCA 1800  
GAAAGAGGAA GAGGAGGGA GGCATTTCT ACGACATCAT CTCTGCGCA CTGCTGCTCC 1860

TTATTTTAC TTGTCTTGC ATGTGCTGT ATTATACA GTTCTGTGTT AACACTTTT 1920  
CAAGTATTTG GGGATTTTAT CTGGCAATCC TCCCTTCTG GTTCTCTGCA CCGACTGTG 1980  
CACTGCACT TCCTTCCGTG CTCGTGACT TTAAAGAGAG AAGGGGGAG GGTCTCCGA 2040  
TTTATCTTT GTTCTTTTTT TCTCTTAGC AGTAGGACTT GATATTTTCA ATTTTGGAG 2100  
AACTAAAGA TGAATAACT GGGTTTTTTT TTTGTGTTGT TTTGTAAAA AAAAAAAAAA 2160  
AAAAAAAAA AAAAAAAAAA AAAAAAAA AAAAA 2196

(2) INFORMATION FOR SEQ ID NO: 164:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1945 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

GCACAGATC GGGGGACCG ACAGGAGAG CAGAGAGGG GTCTGCGCG CGGGCGCTAC 60  
CCAGAGCCA GCGAGCGCCA GCACGAGTG GCTGTGCCC GAGCGCAGCC CCGAGCGAGC 120  
CCCCCCCCG CCGCGNAGG AGGCGCTTC CAGCGAGCCC GACTTCTAGG AGGAGGGGAG 180  
GGGGAAAGC AGCTCAAGCC TCACCCAGCG CCGTGCCTCC AGCCCCGCCA CTCGAGGCT 240  
CCTGTGGACT CGCGGGTCC TCTGTGGAT CTGCGAGGG AGCGGCTGTG CAGAGCGCAT 300  
GGATTTGGTG CTGTCTTCC TCTGCACTT GCTGGCCCC ATGTGCTCTG CCAGTGCAGC 360  
TCGAAAGGAG AAGGAATGG ACCCTTTTCA TTATGATTTAC CAGACCTCTCA GCAATGGGGG 420  
ACTGTGTTTC GCTGTGCTCC TCTGTGCTGT TGGATCTTC CTATCTCTTA GTTCGAGGTG 480  
CAATGTCACT TTCAATCAGA AGCCCCGGCG CCCAGGAGAT GAGGAGGCC AGGTGGAGAA 540  
CCTCATACC GCGAATGCA CAGAGCCCCA GAAGAGCAG AACTGAAGTG CAGGCATCAG 600  
GTGAGGCTT CTGAGACTTG AGGGGGCTG TTGAACCTTT GATGCAAT GTTCGATGCTT 660  
AAGAAACCG GCGACTTCAG CACAGCGCTT TTCCCCAGA GAGCCNAGA ACTGTGTGT 720  
CCCCGACCT ATCCCTCTTA ACACCATTC TCCACCTGAT GATGCAACTA ACACCTGCTT 780  
CCCCACTGCA GCTTGGGTC CTGCGCACT CCGGTGATGT GTGTGTGTGT GTGTGTGTGT 840  
GTGAGTGT GTTGTGCTA ACTGTGCTT TTGTGCTAC TTGTGTTG TGATGTATGT 900  
GTTTGTTAGT GAATGTGGA CTCGCTTTC CAGGCAAGGG CTGAGCCACA TCGCCATCTG 960  
CTCTCTCCG CCCCCGTGG CTTCTCATAC CTTTGTGCTC TAGGAGGCTG CTTGTGTGCC 1020



GAAGACAGACC CCGTCCCTCTG ATTGAGAGAT GCGTAGAGTA AAGACAGGG CAGGTGCTT 1080  
CAGTGTCTT GGAAGCTGGG AAGGTGTGA GCACTTGTG ATCAATCTTC ATGAGACTCT 1140  
TTGACTGCTT TAAAGAAAAC CTTCCTCTCT TATCCGACTT GATCCGATG TGAAGGTCTC 1200  
TTAGACACTG GAGATAGAAA GCAAGAGACT GGTAGAGCCA GCGTAGACT CAGAGAGGCT 1260  
ATGCGCTTCC GTGTGTAAAT TCTTCCAGAG GCGTCCAGG AAGAGTCCC ATTCGCCCC 1320  
CCCTTTCACA GAGCGCCCCG GATTTCCAGG CCGAGAGGCT CTAGCTGAC CCTGAGGAAAT 1380  
GTGTCCCTCT GATATCTCTT CAGCAATAGC TCCATGAGCT CTGAGAGCTT ACCCTTTACA 1440  
ACCTTCCCTG CTTCAGAGAC TTGATATAGC ACGCCAGCTC ATCCAGATGC AAGCTAGAGT 1500  
CCCTGCAATT GGTCTCTGCG CAGCGAATAG TTGAGAGACT CCGTGTCCCT TGGGCGCAGAC 1560  
AAGCCGGGAT GAGTGGAGGG AAGACAGAG CCGTTCCTTC TCTGCCTAGG TCCCTTTAGAA 1620  
TGGGACAGAG AAGCACTCC CCGATCTCTT GGTGTGCTCG TGGTGTATCA GAGCGGTAGG 1680  
CGAGTGTGAT TGGAGACTCA GAGAGCTCC TCCAGCCCTT GCGAGAGTGG AAGGTGTGAA 1740  
GCTGTATAGG AAGTGTGAAA CTGAGCCAG ATCCGAGCCC TCTGTCTCTC TGTGTTCGCG 1800  
CGAAGACAA CGAAGCGTGG GCGTGTGAC CATTGCTTCT CTCTGTATGG TGAATCTATC 1860  
TCAAGACAA CAGAAAAAG GATTAAATTA TCGTTGTCTT CCGTAGTAAA AAAAAAATA 1920  
AAAAAATA AAAAAAATA CTGCA 1945

35 (2) INFORMATION FOR SEQ ID NO: 165:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2933 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

45 GAGTGAACC ACGATCCGG GAGCTGCTT TTGAGTCTT GCTTCCCTCG CCGCTTCCCG 60  
GATTCAGAGC CAGTGTATAC CCGCCGCCCA CCGCTTGTGT GCGCGTAGAG GAAAGAGAAA 120  
GAGAGCGCGC TCGGTCTTGT GCGCGGACT CCGCCAGCTG CTGAGAGAG CAGAGCCCCG 180  
CCGAGTGTGT CCGTTCACAC CCGCTGTAGC GCGCCGCGAC ACGCCGCCGC 240  
CGAGAGAGCG GCGCGCATAG GCTTCTGAGG CCGATTCAAA AAGTGTATAG CTATCAGAGG 300  
CGATTCTGAA AAGAGAGAGC CCGCCGATC GGTATATGT TGAATAGAGC ATCAATAGAG 360  
ACAAGAGTGT GGTGTCTTGG TCCAGGCCCA AAGTGTATGA ATTGCACTTG TTCCGAGGTG 420  
ACAAGAGTGT GGTGTATGAA AAGAGAGAGC GAGAGAGCTT TTGCAATCTC CTTCCTAGAG 480

AAATCTCTTC TGAATGAGAG ATTGAGATGA ATGAGAGTGT TCGAGATAGC CTTCCTATAC 540  
GCGTAGAGAG TGTGATCAGC ATCCAGCAT GCGCTGATCT GAGGTAGAGC AAGCGTATCC 600  
AATGCTGAGC CATTGTATAGC ACGTGTAGAG GCAATATCTG TATATCTCTC GAGGTATAGC 660  
TTAGAGCGTA CTTCCTGAGG GCGTATGAGC CAGTCCGAAA AAGAGAGATT TTCTCTGTCC 720  
GTGTGTGAAAT GCGTCTCTTG GAGTTCAGAG TGGTGTAGAG AAGTCTTAGC CCGTATTTGCA 780  
TTGTGTCTCC AAGACAGATG ATCCACTGCG AAGGAGAGCC TATTCAGACAA GAGAGTAGAG 840  
AAGAGTCTCT GATTGAGAGTA GGTGTATAGT ACAATGTGTG CTGAGAGAG CAGCTAGCTC 900  
AAGTAAAGGA GATGTGTGAA CTGCCCCGTA GACATCTGCG CCGTCTTTAG GCAATTTGTG 960  
TGAAGCTCTC TGAAGAGATC CTGCTTTAGG GACCTCTGAG AAGAGAGAG AGCCTGATTTG 1020  
CTGAGAGCTT AAGCAATAGAG ACTGAGAGCT TCTTCTCTCT GATCAATGAT CCGTAGATCA 1080  
TGAAGCAATT GCGTGTGTAG TTGAGAGACA ACCTGTGTAA AGCCTTTGAG GAGCTGTGAA 1140  
AAGATGCTCC TCGCATATC TTGATGTATG ACGTATATGC CATGCTCTCC AAAAAAGAAA 1200  
AAATCTATAG CAGTGTGTAG CCGGCGATTT TATTCAGATTT GTTAGACTTC ATGAGTGGCC 1260  
TTAGAGAGAG GCGATATGAG ATTGTATAG CAGCAAGCAA CAGAGCCAGC AGCATTGAGC 1320  
CAGCTTAGAG GCGATTTGAT CCGTTTACAA GGAAGGTAGG TATTTGAAAT CCGTATCTCA 1380  
CAGAGAGCTT AAGATATCTT CAGATCCATA CGAAGAGAT GAGAGTCCCA GATGATGTGG 1440  
ACCTGAGACA GTAGCCAGATG AAGTATCAGG GCAATGTGAT GCGTAGTTAG CAGCCCTCTG 1500  
CTGAGAGCT GCTCTGAGAG CAGTCCGAAA GAAAGTGTAT CTGATTGAGC TGAAGAGATCA 1560  
GACCAATGAT GCGAGAGTCA TGAATCTCTT AAGCGTATCT ATGGAATGACT TCCGATGGGC 1620  
CTTGAAGCAG AATTAACCAT CAGCACTGCG GAAAGCTGCG GTTAGAGTCC CAGAGGTAGC 1680  
CTGAGAGAGC ATCGAGAGCC TGAAGAGATG CAAAGCTGAG CTAGAGAGAG TGGTTCAGTA 1740  
TCTGTGTGAG CAGCCAGACA AATTCCTGAA GTTTGGGATG AAGCTTCA AAGAGATCTT 1800  
GTTCATATGA CCTGCTGAGT GTGAGAAAC TTGTGTGTGAC AAGAGCATTG CTATATATAG 1860  
CGAGCCAGAC TTGATCTCA TGAAGGTCC TGAAGCTCTC ACGATGTGAT TTGGGAGATC 1920  
TGAAGCGAAT GTGAGAGAAA TCTTTAGCAA GCGCCGCCAA GCTGCCCCCT GTGTCTATTT 1980  
CTTGTATGAG CTGATTTGTA TTGCAAGGC TCGTGTAGAT AAGATGTAGG ATGTGTGTGG 2040  
GCGTCTGAGC CAGATCTATCA ACGAGATCTT GAGCAGAAAG GATGTGATAT CCGCAAAAAA 2100  
AAATGTCTTC ATTCATTTGCG CTAGCAGAGG GCGTGTAGTC ATTAGTCTCG CCAATCTCAG 2160  
ACGTGAGCGT CTGATCTAGC TCAATCAAT CCGCACTTCTT GATGTAGAT CTCCGTGTTC 2220  
CATCTCAGAG GCGTACTGCG GAGAGTCCC ACGTCCAGAG GATGTGACTT TGAAGTCTCT 2280

5 GCGTAAATG ACTAATGCT TCTCTGGAG TGACCTGACA GAGATTGCG AGCTGCTTG 2340  
CAAGCTGGCC ATCCCTGAAT CCATCGAGAG TGAGATTAGG CGAGAACGAG AGAGCGAGAC 2400  
AAACCCATCA GCGATGAGG TAGAAGAGA TGATCGAGTG CCTGAGATCC GTTCAGATCA 2460  
CTTTGAGAA GCGATGCGCT TTGCGCGCG TTCTGTCTAGT GACATGACA TTGCGAATTA 2520  
TGAGATGTTT GCGCCAGACC TTCCAGAGAG TCGCGGCTTT GCGAGCTTCA GATTGCTTTC 2580  
AGGAACGAG GGTGAGAGTG GCGCCAGTCA GCGCAGTGA GCGCGCAGAG GTGCGAGTGT 2640  
ATACACAGAA GACATGATG ATGACTGTGA TGCGTAAATG GTGCTGGCCA GCGTGCAGTG 2700  
AGCTGCGCTG CCTGGACCTT GTTCCCTGGG GGTGCGGCGG GTTCCCGCAG AGAGGAGCA 2760  
GGGTGCGCC CACAGCGCTG TCCATTCTCC AGTCTGAA CA GTTCAGCTAC AGTCTGACTC 2820  
TGACAGGGGG GTTCTGTGTT GAAATATACA AAACAAAGCG GATTAATATA AGCGATTTT 2880  
CATTTGTTAA AAAAAAAAAA AAAAAAAAAA CCGCGGCGGG GCGCGAGCA TTT 2933

25

(2) INFORMATION FOR SEQ ID NO: 166:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2243 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

55 TCGAGAGCC GCGCGCGG GCGCTCTGCG CCGAGAGCG CCGCTGGAG GCGGTGAC 60  
GATGCGCAGA AGTGGCGTTG GCGTGGGAT CACCATAGCT TTCTGAGTA CCGTCAATAC 120  
CGAGTTTCTC GTGTATATG GTGTCTATCA GTATACATCC CCGATTCTCC TCTATATTCG 180  
TTCTTGCTC CCGTGTATAT TTCTTCAAG AGCGTCAAG GTGCGGAA TAGAGAGACA 240  
GTATAGTAG GTGTGCTG AAAGCGCCA TAGTGATTGA GTCTTCAAAA CCAAGCTTC 300  
TGAGAGCAG GAGATTTTGG GAAGAAATC TGACTGTGGA TTATGACAAA GATTATCTTT 360  
TTCTTAAAGT AATCTATTA GATGCGGCTG ACTGTACAA TGACTCTCTG AAAAACTCT 420  
TCACCTAGTC TAGATAGGG AGGTGAGAA TGATGACTTA CCGTGAAGTC TTCCCTTGAC 480  
TGCGCGCACT GCGCGCTGTC TGTCGCTGCG AGCATCTGCG CAGGCTAGG TCGGTTCAAG 540  
CAGGTGCGAG CTTCGCAAGT ATTGATTTTC ATTCAATGTA TTAAACAG TTGCAATATT 600  
TCAAAGCCTT GACTAGAGC TCATATACA ACCGCAAGT TTGTGTAGT GCGCAAGGA 660  
GATAGTTGA TGCTGCTTAA CAACATGAA GTATGTGTA ATAGGATTA TATTTATCA 720

60

5 AAGATTTTT AAAAAAGGG CTGTGTTTA AAAAAAAG AAAACAGGA AACACAGT 780  
GATTATAG AGCTGCACT CTAAAGTGGG TCGCGCGTG GCGAGCTTC AGGTCAGCG 840  
TCTGCTGTC TCGAGTGGG TGTTTACATG GTACACATG TGTGTATCAC CAGTGGGTCA 900  
ACTGCTGTC ATTCTGCGG TCGCAGTTTG TGTAGCAAT CTTACTGAGC AAAGGCAAT 960  
GAAAGCTTT GGTTCGACA CTGCGAATA TTGGAATTT CACTCAGTT TATGAAGTTT 1020  
ATTTCGAAT CCAATGCTAT CTAGAGTGA ATACTGTCT GCGATGTATT TCATCTTAG 1080  
TGAGCAAAA TTGTGTTT GTTACTACAG AATAGAGATG ACTGTTTTTT GCCACAGCC 1140  
TATGGAATTT GCAATCTGTG ATTGCTTGT AAAAAAGGA GTGCATATG CACTGCATTA 1200  
AAGCTGTGT GTTCTAGTC AATGATATTG GTGAGACAA TGTATCTATT TATGCGATA 1260  
GACCATACA GACATAATTT GCAAGTATTG GGTCTTAAC TTCAAGTGA ATCTATATGA 1320  
AAACCAATCT GAGCTTGTA TCTCTTAAT ATTTATTTTT TTAACTGT GAGATGTTG 1380  
AGAGAGCTT TCCATTCAT TTGAGTGTG CCGGAGGAA ACTCGGCAAT GATTCTTTC 1440  
AGTTGGAAG TTCTTGTGT GTTACAGCTT CCACTGACC CTCAGCTTC GAATACTGC 1500  
AGTTTGTG GTTGTGCTAT TTCTACTTAT AAATTTACCT TTTGTATT TGCATTTAC 1560  
AGCTGTTTG TTGTTTAAA ATTCTGCAA AGTGGCTTGA TTAAAGACT CTTTAAAT 1620  
GGAGCCACC AGTCAGCAGA ATGAGCTT AGAGGACTT GCGTGTGAGC GCTGTCTTT 1680  
GTGTTGTT TTGTGATGA AGCATCTTG CTGCGGTTTT TTGCTTTGTT TTGAGCGAAA 1740  
TGCTGTGAG TAAATTTAA GTTCTGGAG TTAAATTTGT TTACAGGAAT TTGTTTTTT 1800  
AAAAATAG GATCATCTG AACTTTGGA TCAGCCCTTT ATATATTTTC TGAATATGA 1860  
AACAGTTACA TGAATAAAT TTCAATGAA GATGTCAGA TTTTATGAAA AACAGAGT 1920  
TATAGATGA AGCAGCGAG TGATCTTTTA AAMGAGCTT GATCAGGAC ACACATTAAG 1980  
TCTTCTCTC GGAACCGGA AGTAATCTA TATCTGTAG AATTAATGTA GCGAAGAAA 2040  
TGTAAATTT AGATTTTTTT TGCCAATAGT TTATAGAAA TATATGACC AAGTGATTT 2100  
GAGTTGTAA AATGTAAA TAGTATGAC AAATTTGCA CTCTACGAG TTGGAACATC 2160  
TAGTGAAGT CAGTTCTA CTAGTTTTT ACATGTTGT TCTTTTGA TCTATTTTTT 2220  
ACTTTTATA AAGGTTCAAA ACC 2243

(2) INFORMATION FOR SEQ ID NO: 167:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1816 base pairs  
(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

5 GGTGGGAGAC TTGAAATTTT CCTTAAKMG GGGGAAACTT TCCCGAATT  
TTGGGTGCA CCCAGCGTC GGGCAAGCT AGGAGAGAA GTTGTGACT CAGAGGTGA 120  
GCCAGAGACC GAGATTCTT GGGGGTGA GCGGAGCTG AAGTGAAGC GAGGGGAAA 180  
GAAAGCTGG GTGGGGGGA GAGGGGGCT GGGCCAGAA GAGAGAGAA ACCCTTCCA 240  
GAAAGACA ACAAAGTGA GTGTGTGAC AAGAGGAGAC AAGATGGCG GCGCGAGGG 300  
GAGCTCTGG GTGAGAGACC AACTGGGGCT CCGCGCGCTG CTGTGTGCA CGATGGCTT 360  
GGCGGAGGT TGGGGAGCG GTTGGGCTGA AGCAATTGAC TCGATCTTGG GTGATAGGC 420  
GTCTTCCAC CCGGCTCTG AGTTGACTTA CCGCTTGAC ACCTACCTTA AGGAGAGAA 480  
GTGTAGCCA TGTGAGAGG GTTGAAGCT GTTTTCAATT TGTGATTTG TGAATGATG 540  
AATTGACTTA AATGAGACTA AATTGAAATG TGAATCTGA TGTAGAGAG CATATTCCA 600  
ATCTGATGA GATATGCTT GCAATTTGG KTCGAGAT CATGCGAT TCGCTGACT 660  
GAGAGAGAA GACTTATGT CCGTGAAGCC AAAAATGAC CTAGCTTTC GTTAAACT 720  
GGTGAAGTCA TTTGAGATG ACATGATGA CTCGAGAGG AGCTTCAATA CCGTCTGAT 780  
GACTTTTAT GTTGAAGCG ATGAGAGAA AATGATTA TTTGAGCTA ACCCGAGAA 840  
TCCCAAGTAC GAGAGAGAT TGAAGCGAG AGCGCTACA AATTGAGAG GAGCTCTCT 900  
AAGCAAAATG TCCATCAAT GCGAATGAG AATTTACAA GCGAGAGAA ATTCTTGA 960  
AGATGAGAA AGTGAAGCT TTTTAAATG CCGTCTCTT AACTGTGGT GATTTTAC 1020  
TACAGCTCT GTCTCTCGG TGAATGATT GCTTGAAT TGTGTGCA CTGTGTGCTA 1080  
CAAGCTTGG GAGAGATAT AGTTTCCCT TGAAGAGTG AATATCTAG GTGACTTGA 1140  
GTTTGAATAT GAAAGAAAG TAAAGAGTA TCCAGCTCT TCTCTTGGG TTTTGAATC 1200  
TAAAGCTGA GATCATGAG AGCAGAGCC TCTACCTACA AAGTGAATC TTGTCTATC 1260  
TGAATTTAA GCAATTTCT TTTTAAAGAC AAGTGAATA GACATTTAA ATTCACTCC 1320  
TCAATGAGCT TTTTAAATGG TTTCAATGA TATAGAGCTT AAGAAATAC TATTAATGC 1380  
AATTAAGTT ACTCAATCT GTGAGAGAA AAAAAAAG AAAAAAAG TCGAGGGGG 1440  
GCGGTGACC AATCGAGCT AATGAGATG GTATTGAT TTTACTATA TCTGACTTA 1500  
TTTGTTTT KCGTGGGT AATGTTTT TCCCTTTCT AAGCTATAG CTGATCTG 1560  
CTCTCTCA CCGCTGCA TGAATCTG AATTACTTA GTTAAAGAC TGAATATTA 1620  
GTGAAATGA TCGTCTGT CAGAGAGCC CAGAAATCT GTTAAATGA ATTAGAGG 1680

AATGACTT TATGACTT ACATTTGAG GAGTGAAT CATTAATTT TATTTGAT 1740  
AATATGTC TGAAGAGAA AAAAAAAG AATGAGAGK RAAATCTCA GCGCGGCCCC 1800  
GTAACCAAT TCCCGG 1816

(2) INFORMATION FOR SEQ ID NO: 168:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 945 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

20 AAAAACTTT GATGAGACT AGAAACGA GTTAAAGCT GTTGAAGCT TGTGAGACT 60  
CACTTGAC CAACGAGAC AGTGGAC ACATCAACT TGTCCAGAC AGAGAAACC 120  
GAACCAACA ACCAGGGCA GAAAGCTG AAGAAATAC TACAGAGAA AATCAAGTT 180  
ATTGAGACTA TCGAGATTT GTTGAGATG ATGTATTTGA GCTTGGGAT CATTTTCCA 240  
TGTCTCTCT TCTCTCAGT TTTTACCGA GTGACTCTA CAGTTTGA CTCTCTTAC 300  
CGATCTGAG GAGCTTTT TTTTATCAT TCTGCTCTC TATCAATGC CAGAGAGAA 360  
AGTTTACA AGCTTTTGT GCAATGAGC CTGTTTGA GATCTGAG TCGTCTCTT 420  
GCGTGTGG GTTCAATAT CCGTCTCTC AAGAGAGCA CTTAAATCC TCCCTCAGC 480  
CACTGTGAT TGAAGAGAA TAAATTACA ACAAAGATT ATGTTCTTA CTTTATCAT 540  
GATCACTTT ATACAGAGCA CTGATTACA GCGAAAGCA GTCTGAGTG AATCTCTCT 600  
CTGATGCTGA TTTGACTCT GCTGAATTC TCCCTAGCTG TCTCAATGC TGTCTGCGG 660  
TGAAGAGG CTATCTGA CTTCCTGGG AGTGTACTT TCTGCTCA CAGTTCAAT 720  
GTAATCTG GCAATGCTT AAAAAATCT CAGAGCTG GATTTAGAA ACTATGACT 780  
TCTTAAAGT AAGAGAGAA AATTTATCA GAAAGTAT TCTTATGATA AATGAGAAA 840  
GTAAACCTT ATAGAAAGC AAGCTGAG TTTCTTAAT GTTAACTTT AAGTATGTA 900  
ACATTAATA AATCAATAT TTTCACTGA TTTAAGATA ATGTG 945

(2) INFORMATION FOR SEQ ID NO: 169:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 902 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

```

5  GGCAGAGCCA CAGGAAGAGT GAGGAGACC AGGCTCTGGG GGCCTCTGTG GATGCTCTTT 60
   GTCTCAGAAC TCCGAGCTGC AACTAAATTA ACTGAGGAAA AGTATGAACT GAAAGAGGG 120
10 CAGAGCCCTGG ATGTGAAATG TGACTACAGC CTAGAGAACT TTGCCACAGC CCAGAAAGCT 180
   TGGCAGATAA TAGGGAGAGG AGAGATGCCC AAGACCTCTG CATGCACAGA GAGCCCTTCA 240
   AAGAAATCCC ATCGACTCCA AGTGGGAGG ATCATCTAGT AAGACTACCA TGATCTGTGT 300
   TTACTGCGCG TCCGAAATGCT CAACCTTCAA GTGGAAGATT CTGCACTGTA TCAATGTGTG 360
   ATCTACAGC CTCCTCAGGA GCTTCACATG CTGTTCGATC GCATCCGCTT GGTGGTGACC 420
20 AAGGGTTTTT CAGGGACCCC TGGCTCTCAAT GAGAAATCTA CCGAGAAATG GTATGAGTT 480
   CCTCTACCA CCACTANGGC CTGTGCGCA CTCTATACCA GCGCCAGAAC TGTGACCCAA 540
   GCTCCACCCA AGTCAACTGC CGATGTCTCC ACTCTGACT CTGAATATCAA CTTTACAAAT 600
   GTGACAGATA TCATCAGGCT TCGGTGTTC ACATGTGCA TTCTCTGGG TGGTGGATTC 660
   CTGACTAAGA GCTGTGCTTT CTGTGTCTG TTTCGTGCA CCTGAGGTC ATTGTACCC 720
30 TAGGCCCCAGC AACCCACAGC AATGTCTCTT GACTTTCAGC CACATCAATC TGGCAGTTGT 780
   GCGAGGGAG GAGGGAGAG GTAAAGGCA GGGAGTTAAT AACTGTGTA 840
   TCACCCGCTA AAAAAAAAAA AAAAAAAAAA CAGACTGCG TTTCAGCTC CATCAGCTCC 900
   TT 902

(2) INFORMATION FOR SEQ ID NO: 170:

(1) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 1883 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

50 AGAAACACAC TGAANAACCA CATTTTTCTA CATACAGCTG GGGAGGTAGC TGAGAACTTG 60
   GCATCGCCA CACATCTAGT GTTGAAAGAG AGTTGAGAAA ACCAGAGGC CAGGTGCATC 120
55 TCGTGGCMAA CCTCGAACT GTCTCTGGC CTTCCTCTAC AGTCTCTAG TTGAAATACC 180
   TTTTCATGCC TAGCATCTGC TTGAGTTATA AACCCACAGC CAGCATGTC ATAGACTAGT 240
   GTTACTCTTT GTTTGAGTT TGTTTAATG CTTCTTAGA CCAAGTCCC TCCTCTGTT 300

```

```

5  TCTCTCTTGG TGGTAGCCTC TGGCCATCTG GGACCTCAAT CCCAGCTTT CCCACTTTCA 360
   CGAGTCTTTT GCTCTCTTGG CTTCTACTC AATAGCCCC AGGAGTGGC TTATGCTCC 420
   AATATCGAGC ATTTCAAGCT TCCTCTGGG GATGGGANT GGGATGGCCA GAATCTGTTT 480
   TGGTCTCGG GATTATTTCC AGTGGGTGA AAGGAGAGC TGGGCTTTTC CCTCTCTTAT 540
   CCTGAGGGT GGTAAAGAG GACTGTATCT ACACCTGTTT TCCTCTACTT TCTCTTTTGT 600
   TAGGAGGCC TCATCTTAG TTCTTCAGA GAGTCTCTGG CTTAAGCTG TAGCAGGGT 660
   GTCTAGGTG GGGGATTTGG ACCAAACCG TCGAGTAGGC ATGATFAGTG TATGAGGTGG 720
   GCTCTCAAAA TCAGACAGAA ATGGCTTAG AAGCCGAGG GGAGCATGCC TGTCTCTCAG 780
   TCATAGAGTA TGGGAGGAGC CTCCTTAGCT TGGAAATGA GAATTGAGG GGTATGAGC 840
   AATAGAGTG CCTAGTTCAG GATGTTGCCA AAGTTTGTG CAATCTTATC ATTAGTAGAT 900
   TTTATAGCC ACAGACAGAA ACCAGAAAG GAATATGTT ACTTTGAGT CTTTATTTTT 960
   TTGTCTAGG TGTGCTTGG TACATCAGA AGAATGCTAT ATCTCTCACA TTTTGCCTTT 1020
   AAGTCTTAC GACTTTGCCC ATTTAGTCT AATGGAGGA TACAGATGTG CAAGTCTGCT 1080
   TTTTGTTTT TTGTATTTAT TTTTTTTTT TTGCTCTGTT TTATGGAAAT TTTAGAAAT 1140
   GCAGAGAGT GGAGAGATG GTCTTGGAC CCAATGTGTC CATCACTTAG CTCATCACT 1200
   TATCAGCTAT GGTCAAGCTG GTTTCATCTG TATCTCTCTC TTTTCACCTG TATTTGTTAT 1260
   TGAATTCCA AGAAGCTATG CCAATCCAC COTGACTACT TTGGGAGATT GGTACTCTCT 1320
   TTTGATGTTG ATAGTATGCG GGTGACTAT CATATCACA TCAGTCTCTC TTTTGTCTTT 1380
   TAAATGTAAC TAATGAAGTT CCAGAGATGG CCGTTAGAAA TGTGTTTTAA GAATTACAA 1440
   GGAGTCTCAA AAGAAATGA GAGGATGCT TCCTTTGCC TCCTACTAC AAAACAGAG 1500
   AGAGACTGTT CTGTGTAAA ACTCTTTCAA AATTTGTGAT ATGTTAGGT ACTTGAGCC 1560
   CTTACACAGA ATGTCAATCT TTTTTCCTGT GTACATGGA AACTTGTGT ACCNTAGCA 1620
   TTGTTATCAG CTTGTACTGG TCTCATACT CTGTGTTTGG AAGATAATTT TCGAAATGTT 1680
   TGTGTGTTT TGTGAAATA ACTTCCCAA AATATTTAGT AACTGTGTTT TCTACTTGTT 1740
   AATTTGACAC CCGTTAATA AGCAATTTAT TTCTGTGTTT TTAACAGTA TAATAGTTG 1800
   TAAGTTTCCA TGCATGAGG AAAAAAANA ACCTGTATCT CTGTTAANA AAAAAAANA 1860
   AAAAAAAAA AAAAAAAAA AAA 1883

```

(2) INFORMATION FOR SEQ ID NO: 171:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2100 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

5 TACTTTAGA TTACTGCTT TCGAAAAGTG CATTATCGA GCAACATGAA CATTATCTCT 60  
10 TACATATGTA TGTACACAGAG GTACCCAGAG TCGTACTGTG GCAAGCTTCA AAAACATACC 120  
ATCGAAGAA GTAGGTGTCTG AGATTAAGATA ACTTTGCCAA ATGAAAGAAA GTCTCTACT 180  
15 TCGAATATCC CCTCTTCAG CCGCTACCTGT GAAAGGGGCT GCAACACAT TCCCTGAGCA 240  
TCCCTGTCTG ATACAGCTTC TTATATTTTA TATCTTACTG GATGTAGCA TATGTCTAG 300  
GTTCCTGTGA GTCTCTTCA AGGGAATGTA AGTTTATGAG CATTGAAACA TTTCAGAAAA 360  
20 AAAAAAGAT TTAAAGAGAT TAAATAGACC GTATCTGTGA TTAGATGTG TGTCAATATG 420  
GTGTCTATGA AACTATAGAT CAGTGGGTTT AGATGTGTTA AGTGTAGCA CATTCTCTCT 480  
CCTTTTGTCT GTACAGCTGA CATTGAGATA AATGAAAA TGTTGGCTGT GGGGATGTGA 540  
25 AGCTCAGGAG GCGAAATGTC GTTGCAGATC CCTTAGAGCA TTACTTTGAC TCCATAAAAT 600  
AGTATGTAT GTTATTTAT GGGTTTTGT TCCATATGTC CATCTAGAC AAAACTGTCA 660  
30 AATATGTGA TGTAGAGACA GCAATGCTGA GATGTATGCA TTCTTACCCA GAGGTGTCAA 720  
TACAGAGAG TGCATATGAA TACAGAGAG TACAGAGTC ATGATTTGAG GAGAGGGTTT 780  
35 GAAAGGAGGA CATGATTTCA AAAAAAGTGG TTCTCAATGT GTCTGTAGAC TCAACAGACT 840  
GGCAATTAG ACTTGTAGAG TCGTTCCTCT TCGTTCTAG TCGATGGCTT CCAATTTGAC 900  
TTGAATATGC GTCTGTGAG GCAAGAGAG AATCTCTGAC TTAACTTTGA TCCAGAGAG 960  
40 GTCTGTGTGT CCTGTGTGTC AATAAATGT GTCTTACTTA ACCGAGTTTT ACCAAATGGA 1020  
AATAAAAAGG GACAAACTAT GGAATATGTA GTCAATGCA TTGCATGTAG CCAACATCTT 1080  
45 GTTTTCATTA TTAGAGACC CATTACATTA GCTACAGGTTG AGGTGTAGAC AGCTATGTTT 1140  
CAATATTTCA AGATGTGAC GACCTGTCT TATGTATCAT CATGTATGA ATTCAGTTGA 1200  
CTCTTTGCA AAGGGTGTAT ACTTTTCACT AAAAATGCTT ACTGTCTG TGTATGTTCC 1260  
50 TTTTGTGTT TTACTGTGTC CATTTCGAC ACTATGCAAT TTTTATTTT TTTAGAGAT 1320  
CAATTTTAG CAGTGAATA TGAATGAAA AATTTCAGTG TAAATGTATA AGAATGTGAG 1380  
GATACAGGA TTTTTTTTTT CCTGTGAAC AATGTAGCTG GAAAGAAA CACATGTCT 1440  
15 TTCTCTGAG TTTCATCTG ATGATATGA CCAATGAGGA TATCTTAGG TAAAGGTTTA 1500  
ATGATGTATA CAATGTATGA GATTAAGGAG AGATGTGTAG AAGCCGGGTT TTCTCTATGC 1560  
60 TAAATGTGTC TACTTAGAGC AGCACTTCTT ACTATCTAG CACATGTATA GCGCCACTCT 1620

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1930 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 172:

20 CCTTGAATG TGTTCCTGG TCGAATATG CAGCGCTGCC GCGCGGCTCT GTGTGTGTC 60  
30 CCGTGAATG TGTTCCTGG TCGAATATG CAGCGCTGCC GCGCGGCTCT GTGTGTGTC 120  
35 GGTGTGTG TGTTCCTGG TCGAATATG CAGCGCTGCC GCGCGGCTCT GTGTGTGTC 180  
CGTTTCTGC CAGATGACG TGTATGATCT GATATGATCT GATGTAGCA ACTCTAGAG 240  
40 AGATTTTGG TTTCAGAGCA AAGATTTGG AAGATTTGG GCAAGATGTC CCAAGAGAG 300  
TGTATGAGG TTTCAGAGCA AAGATTTGG AAGATTTGG GCAAGATGTC CCAAGAGAG 360  
45 TGTATGATCT GTGTATTTT TCGAGAGTT TATCTGCTT ATGATGCTT TGAATAGAG 420  
AGTTCGCA AAGATTTGA AATGATGTG CTTGAGACT GATTTTCTCT TTATGAGAG 480  
45 TGTATGATCT GTGTATTTT TCGAGAGTT TATCTGCTT ATGATGCTT TGAATAGAG 540  
TGTATGATCT GTGTATTTT TCGAGAGTT TATCTGCTT ATGATGCTT TGAATAGAG 600  
50 TGTATGATCT GTGTATTTT TCGAGAGTT TATCTGCTT ATGATGCTT TGAATAGAG 660  
AAGATGCT GATGAGAG TATCTGCTT CAGAGTAC CTTGAGACT CTTGAGACT 720  
55 AAGAGAGC GAGGAGAG TTAGAGATC TAAAGAGAT GGTGAGATTA AAGATTTTC 780  
TACTAGAG TGGCTTAAT CACTGAGTA TATATGATC AACTTTTGA CCAAGCTTTT 840  
60 TAAATGATC ATGATGATC TGTATGATC GACAACTTA TGGCGAATA GAAAGCTTCC 900

AGTTCCGTTG GACTGGCTG AGTACAAAG TCAGGAGAA GAAAGCATG CATCAGTCA 960  
ACAGAAATGA CCCCACTTAG GCTTAAAGA CAGCAGGTT CTAGATGTA AGACTATGC 1020  
5 ACCTCTTTT TCAGAGCA TCAGACTTT GAGATTTAT TTAGCAGAA AGGGGATGG 1080  
AGCTGAGTC ATATGGATA AGATGAGCC ATCTGCANT GATTTTGCA CTTCTGCTGC 1140  
10 AAAGCTCAG ATGCATTAAT TCAATATGA TATGAGAGT AGATTGATA TCANATCAT 1200  
GGCAGGAGC ATTAATTTCTG CTATTTGAC TACTAATGCA GTAAITCTG GGTGTAGT 1260  
ATTGCAAGG TTGAGATTT TATCAGGAA ATTAGAGCAG TCAGAGCAA TTTTITTTGA 1320  
15 TAAACAGCA AACCCAGAA AGAGCTTCT TGTGCTTGT GCACTGATC CTGCCAGCC 1380  
CAATGTTAT GTATGTCCA GAAAGCAGA GGTGACTGT GGGCTGAATG TCATTAAGT 1440  
GACTGTTCT ACCTTACAG ACAGATAGT GAAGAGAAA TTGTCTATGG TAGCAGCAGA 1500  
20 TGTCAAAAT GAAGATGGA AAGGAAAT CTAATATCT TCCGAGAGG GAGAGAGCA 1560  
AGCTAATAT CAGAGAAAT TGTCAAAAT TGGAATTAGA AATGGAGCC GCTTCAGCC 1620  
25 AGATGACTTC CTCAGGACT ATACTTTAT GATCAGATC CTTCATAGT AAGACTAGG 1680  
AAGGAGCTT GAATTTGAG TTGTGTGTA TGCCCGGAA AAGTGGGGS CCAGCAGCC 1740  
TGAGATGCT GCGAAAGCA TACCAATGG GCAGTGAAGA TGAGAGCTTC AGCCCTCCAC 1800  
30 CTTCAGCT TCAGGAGCC AGATGGAGG TTTCATAG TTGATTCGBR TGAGAAAGT 1860  
TCTCAATAA TTGCCGAGG TTCAITGAG GAAGGAGAG GAGGCGGCC AAGAGGGGA 1920  
35 TTAGGNTTG 1930

(2) INFORMATION FOR SEQ ID NO: 173:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1509 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

50 GGGCTTGGCC TCTGGAGTGA GCTTGTGAG GCACTCGGG TGCTCTTAG GGGCAGGAT 60  
AGGGCTGGGG AGCGCGGCC TGTGGCCGT ACCAGCCCT TCTGTCTG GTTCCAGCC 120  
GATGCAAGTG CTCAGTCTCT TGACGGGGA CAGCTACTTG AGCACTGCT TCTCCAGCA 180  
55 CTTCTAGTC GTGCTGCTCT CTCGCAAG CAGCGCTCG CCGAGGCTG TTGACAGGA 240  
CTTCTACTCC GAGTTTGGG ACGAGACC AGGAGAGTG GAGAACTAG AGCTGATCA 300  
60 CTCATGTCG CTCAGTTTA CTTACCCAG TCAGAGAGAG ATTGGGAGC TGACCTTAC 360

5 TGTGGCCAA AAGATGGCTG AGCCAGAGAA GGGCCAGCC CTCAGCATCC TCTGTAGCT 420  
GGAAGCTTTC CAGTGGGCA TCCACCCCC TGGGTGCTGC AGGGGCCCC TGGCCCCAA 480  
GACATCTCTG CTCACAGCT CCGAGATCTT CTTCTGTGAT GAGGACTGTG TCCACTACCC 540  
ACTGGCCGAG TTTCGCCAAG ACCGGCCGA GAGAGAGAG TACCGGCTGG AGGATGGCG 600  
10 CCGGTTCCG GACCTGAGC GAGTGTCTAT GGGCTACAG ACCTAGCCG AGGCTCAGC 660  
CTGTCTTCG ATGAGTGTGA AGGTCAATGAC CTCATGGCA GTGTCACTCT GACCACTTT 720  
GGGAGGTGC CAGTGGCCC GGTAGAGCC AGCCAGGGCC GTGAAGTCCA GTGGCAGTG 780  
15 TTTGTCCCA GTGCTGAGAG CAGAGAGAG CTCATCTGCG TGTGTGCTCG CAGTGGGAG 840  
GGCTGTGTC GCGGTGAGCT GCTGTGCGAG CTCAGCGCT AGCCAGAGCC ACAGCCAGCC 900  
20 TGTGTGTC AGCTCAGCC CTCAGTGGCC AGGGCAGAG GCTTTTGTGT TCTCTAATA 960  
TTTTTATCC TCCCTTTGCT ACCTTAATTT GACTGTCTC GCAGAGATG TGACATGTG 1020  
TGTGTGTGT GTTAATCTT TCTATGTTG GAGTGTGAA TGGCGGGCC CTCAGGGCTG 1080  
25 TGGGTGCT GTACGCTCC CACAGTGTGT ACGAGCGTGC ACAGAGTGT CGTGTCTGCT 1140  
GTGTGGGAC GGTGTTAAC AGCTGACACT GTGGGTGTGA CTTTCTCTTC TACAGTCTCT 1200  
30 TTCTGAACT GTGAGTCCA GTCTTTGTT GCTGTGCTG TTGCTGTGTC TGTGCTGTT 1260  
GGCATCTTC TGTATATCT GAGGCTGTA GCAGATGCA CATTTGAAAG TCCACATCCA 1320  
TATGTGTTT CAAGTGGAG GTCTCCCTG ATCCAGAA GTGGGAGAG CCGTGGGGG 1380  
35 AGGGAGCTG GAGTGGCAG CACAAAGCT GATTCCTGCT GCTGTATTC TCTATTCCA 1440  
TAAGCAGAG TTGACAGC TCATAAAAA AAAAAAAAAA AATTGTCGG 1500  
40 CCTCAAGG 1509

(2) INFORMATION FOR SEQ ID NO: 174:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3173 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

55 TGAGACCCAS GGTTCGTGC TTTCGACAG AAGGTAGAC CTCGAAAGG ATGGCTCAGC 60  
ACCACTATG GATCTGTCTC CTTTGGCTGC AAGCTGGCC GAGAGAGCT GGAAGAGCT 120  
60 CAGAAATCTT CACAGTGAAT GCGATCTCG CAGAGTCACT CACTTTCTCT GTAAATATCC 180



5 COACATGGA GACTCTGAG ATCATTAAGC CCTTAGATGT GTGCTGGTGG ACCAAGAAC 300  
TCTGTGGGTT CTAGCTGGAC AGGGTGTTCAG AGGATCATCA GGAGCCAAAC CCCAAATCT 360  
TGAGAAAT CAGCAGCAAT GCCACTCTT TCTCTACAT GCAGAAACT CTGGCGAAT 420  
GTACAGAA GAGCAGTGT CACTCAGGC AGGAAGCCAC CAATGCCAC AGATCATCC 480  
10 ATGACAACTA TGTATGAGT GAGGTCCAG CTCTGCCAT TAAATCCCTG GGAGAGCTGG 540  
ACGTCTTCTT AGCTTGAT ATTAGATAT ATGAGTAT GTCTGAGT TGATGACAG 600  
GAACCTGTAT AGTATCCAG GCATGAAAC CCCCTGTGGC GTTACTGTG GGAGACACC 660  
15 CACTTGAG GGAGAGGA TGGGAAGGC CCTTTCAGC TGAAGTCC ACTGCTGGC 720  
CTCAGCTGT CTATTTCCG TTGAATAG CCAGAAATG TACTGTGTA TTTGTATTA 780  
20 ACTATATCT CTGAAGGC CTGAGGCCA TCTGGGAGT AAGGGCTGC CTTCATCT 840  
AATTTATGT GAGTCATAT AGTCCATGT TGTATGTCA GCAAGTAT ATCTGTAT 900  
ACATATGTA CTGATGTGTT TTTCTGATA ATTTCTAT TTTACTTAA AAAAAAAA 960  
25 AAAAACTGA GGGGGGGCC GTACCATTT T 991

30 (2) INFORMATION FOR SEQ ID NO: 176:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

40 ACGCCCTCT TGGAGCTG AGCCGGCTC TCTCTACTA CTTCAACCC CAGGGGGCC 60  
CTCCACAGG CCCCTCTCT CCTTGGAGG CTCTGCTGGT CTCCCGCTCC CCTGGAGAG 120  
45 AACAGGCCA TGGGTGGGCC CTTGCTGCTG CCCCTCTCTC YCTGTCTGCM GGCCGACGA 180  
TTCTGACGC CTGTGTGCTC CACAGATCT GTCTGAGCT ACTTTATGG GTTCTCTAA 240  
50 CCAAAACAC TCTAGCCTC CATGGGTGGC TCTGTGGAAA TCCCTCTCTC CTTCATATAC 300  
CCCTGGAGT TAGCCAVAG TCCCTACATG AGAATATCT GGAGAGGGG CCACCTCCAC 360  
GGCGATGCT TGTACAGAC AAGCGGGCT TCAATTCACA AGAATATGT GAACGGGCTC 420  
55 TTTCTGAGT GGAGAGGG TGAAGGAGC GGTTCCTCA GATCTCAA CCTTGGAG 480  
GAGGAGAGT CTGTGTATT CTCCGATC GAGCTGACA CCGGAGATC AGGGAGGAG 540  
60 CAGTTGAGT CATTAGGG GACCAATCT ACCATCAC AGGCTGTAC AACCAACCC 600

5 ACCTGGAGC CGACAGCAC AACCACTATA GCGGGCTCA GGGTCACAGA AACCAAGGG 660  
CACTAGAT CATGGCACT AGCTCTGAC ACTGCCATCA GGGTTGCATT GGTGTGGCT 720  
GTCTCTAAA CTGTCAATTT GGGACTGTG TGCTCTCTCC TCTGTGTGG AGGAGAGGA 780  
AAGTAGCAG GGCGCAAGC AGTGACTTCT GACCAACAGA GTGTGGGGG AAGGATGTG 840  
10 TATTAAGCCC GGAGGACGTG ATGTGAGACC GCGTGTGTGAG TCTTCCACAC TCGTTCCCA 900  
TTGGCAAGT ACATGGAGAG CACCTGAGG ACCTTTAAA GGCAAGGCG CAAGGCAGAA 960  
GGAGGCTGG TCCCTGATC ACCGACTGCA GGAGAGTTAC CTACAGAGC CTTCTCCAG 1020  
15 GAGCATCCAC ACTGCATCA TATAGGATG AGCTGTGAC TCCACTGAT TAAACCACTG 1080  
GCATTTGGG GCTGTTTAT ATAGAGTGC AAGAGTTCC TTTATCTCC CCAAGGATG 1140  
AAGATCAAT TTTTGTGCT TACATACAC CCGTTTCTC CTGTGCCA TTTTCCATC 1200  
20 TGTATGTTG CTGTCTCTA TGGCAGAGG TTTTGGGAA TAAATAGCT GAAATGTTC 1260  
TGACTTAAA AAAAAAAA AAAAACTCA 1290

(2) INFORMATION FOR SEQ ID NO: 177:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2390 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

60 TGGGGCCCT TTTGGATCT CTGGGTGTT TTGCCAGAG TTACAGGATG TCAAGTGTG 60  
40 GGAGCTGAG ACCCTGCTG TGGACAGTG AAGCTGTTT CAGACAGGT GCTTCCAGAC 120  
ATTTCCAGC TCCAGGAGG AGGCTGGAG CCCCACAGA AGCCACAGA AATCCAAA 180  
AAAAAAGCT GTTTTTTT TTTTCTTT TTATTATGA AACAAAACAA ATGCCCAGG 240  
45 AGAGGCTCC ATGATACCA GAACATCA AGAGTACTT CTACCATTT TATTTCTGTG 300  
TGTTCAGGC AGCATGCCA TAACAGCT AACTACTTA CATTGACTC ATTTCTAGTA 360  
50 ACTGATTT ACAGATAT ACTAGAACG GCATTAATA GTTTAGAAA AGTTAGGTA 420  
AATTGATC CACATCATC AGAAAGTAA CATTTTAAT ATAAAAAGA AAACTTCTT 480  
GGAGCATTA TCCAGTATT AAGGAGAGT GCTACTCTG ATGTGACAAA TTTCTATGT 540  
55 GGTGTTACT CTTTCCCAA AGCTGTGAG AGGGTGAGT GCTCCAAAG AACATCACA 600  
AAAAAACA CAAAAAAA TGTGTTTAC AGTTTATAG CAAGATGACA CTGCCAACA 660  
60 CAAGAGGGG TCTGGATTC ATTTACGCC GGAAGCTGC CCGCTCGGCC TCCAGGGGTC 720



ATTGAGAGTG TTCTGAAATC GAAATTCGAC AGAGAGCTTG TGACTACTCC TGTCCCTCTG 780  
AAAAAGCAT GTTGAAGCT GCGGTACAGG TCTGAGCAT GGGAGATCT AATTGAATCA 840  
CGCGACCTT CTATACAGA AGAAAGGAG GTGACTTCA CCGTCAAGCC GCGAGGAGG 900  
GGGTAGAGA AATCATTAAT CCGTGAAC TCTGAAAAA AACAGATTT GAACTGTGA 960  
GAGAAAGCT TAAAAAGTG AGAGACCA TCGAGTGT GAGTGAACC ACGGTGGCT 1020  
GTGAGGTCA GTGCTCTT TCTAGATGAA AGGAGGAGAG GCGAGCCGAC GCGAGCTCA 1080  
GAGAGACCA GCGGAGAG AGAGCCCGCA GATGCTCC TGTGCGTGC CCGCAGACC 1140  
GGCGGCTCC CGAGGCGTC AGAGAGGAG GACTGACTG CCGTGTGGT GAGGGGGCAT 1200  
TGCAAGGAG GCGCCCGAGC CCGAGGAGC CCGGACTTGA GGTGTAGCTA TGACCGAGC 1260  
CTGTGCTGC AGCAGCTTAC GAGCAGGCT GCGTAGAGC CTGTGACTG TGTGTGTGA 1320  
ATTCCTTAA TTGGTTTAA ATATGCGATT AAGATCTGT TTAGAAAAA CCGTGTGAAA 1380  
GAGGGGTAC TTAAAAAT GGAACCTTC AATTCATTT ATATTTTAT TTATACAAA 1440  
ACTTAATTAA AAGTTTACA AACTGCTGA AAATCAACA AGTGTGAGC TGACAGACA 1500  
TTTAAAAAT GATTAATTAC GAGATCTCC TGATGAGT TCCCTGTCA GTAAAGTAT 1560  
ACGTACATCT GTAAAGGTCA GTGACTCTG AATCAATTT ATGATTTT TTAAATCAC 1620  
GTGTATTAG ATACTAATGA TATGCTCTAT ATGTATCGAG AAATCTGAC AGAAAGCAT 1680  
GGCAGCATA GAGAGAGAC GACAGCAGG CTCCATNCC ACGTCTGAC TGAGAGCTC 1740  
CGCAGCTGA GGTCCGAGG AATGCTGCT TTCACTGATT TCCGAGCTAA CTGTGAGAC 1800  
GCGTAGCAG GAGCAGCCT GCGATCTCC GAGAGGGAAT CCGCTGGGG CCGAGAGCT 1860  
CGTCAAGCC TGGGAGGCG AGAGAGCTC GAGAGGCT GCGCAGGCA CTGAGAGCTG 1920  
GAGAGAGCA GCGATTTCA CCGGAGACC TGGAGGCTA ACTGTGCAAT TCCGTGCGG 1980  
AGCGGCAATG CTGATTGCTG GCGGATTTTG GAGATCCCC GTACTGAAAAG AACATGCTC 2040  
AGCGCTGGG AAAACAAAAC CAAAGATCA CCGTCTATTA AACTGTGAT ATTATTAATT 2100  
TTTACATNG AAGTTTAAA ATCAAGACT AGATTACTA TACATTTT CTCTGATTT 2160  
ACAAATTTA TTATTATTA CTGGGCTCC CTAAATTAAT TTCTTTTAA ACAGCTTTAA 2220  
AGAGACCAAG AATGAATACA AAAGACTAA AGAAATTAA AATTAGATTT GTCTGTGAC 2280  
TGAAAGCTT 2290

(2) INFORMATION FOR SEQ ID NO: 178:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 549 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(%I) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

GCAGAGGC ATGCTGAGC TGTCTGAT TTCTGAGTC ACTGTGTG CTGTGTGTA 60  
CTGAGAGCT AACTGCAATG TGGCGAAG ATGACCAATT CTTTTCAG AGGCGAAAA 120  
TGTGATNG TGCTGTCA TGCTCTCT CATGGCTAC GACTCTCC ACGTGTGTA 180  
ATCATTTCA ACCAGGAG AGCTGCTG AACTGACTC TGGAAATCC CTGGATGCT 240  
TTGTGAGG AATGATNG GCAATAGCT GTTTCCTTG ATCTGGGCC TCTTGATGTG 300  
AGTGAAGAG TAAAAAGCA CCAATCTCT GACTCTGAG GAACTGATC ACAAGAGA 360  
TGTTCGAG ATGCTGTCA AGATTGCTA AAATTAAGC GTTTCAGACC CCGTGAATC 420  
ATCATTTCA GATGCTCT TGACGAGAC GAGAAAGTG ATTACAGAA GTGATCTAA 480  
AAGATTCAT CCGGATTT GAGTACTIC AATTATGAT TCTACTATT AAAAAAAA 540  
AAAAAAA 549

(2) INFORMATION FOR SEQ ID NO: 179:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1509 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(%I) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

GGCAGAGG CTGATTAAT CCGACCGGG CCGTCAGAC ACTGAGGCC TTCTGAGCC 60  
GCGGCGGCA TCGGCGGCG GAGCCCGAG CATGTGAGC CGAGAGCTG AAAGCGCTC 120  
GGCATTCAG ATCTGCGGA TCAATGGGCT AATATATCC AACTGTGCG GATATTTCA 180  
CGGGGAGAC ATCTGAGA TGAATGAGA GCGAGGCGC ATCATAGCA CCGGCAATG 240  
CAAGAGCAG AACGGAGAC GCTGTGAGC GAGCTGCT CTGTGAGC GAGAGACTT 300  
CGTGTGCC ATGTGATNG GTGAGGTGC GATGTGAGC GCGAGATCA CTGAGCTC 360  
GAGCAGCT GTGAGGTGC AGTTCAGCT GATTTGAAA AACATCTCA GAGTGTGAA 420  
AAGCTGAGC AATTAAGCA CCGTGTGTA TGTGCGCTG TGGTGAAGA ATGTGAGAA 480  
GCTCTGAG GTGCTCTG TTGTATTC CCGCAGAG GAGAGAGAG AGGCGGAAA 540  
GGGTATGAA GCGGAGAG TGAAGGAT GAGAGAGAG TGAAGAGC GAGATGCT 600

COAGCCAGTC CTCACCCAG AGCCAGACAC TGTCACTAC AGCCAGTCCA CTTTGATCCA 660  
CTGTGTGGG CTTTCAGACT GACCTTCGA CGGCTTTTGT CACGAGGTG TCACATGAA 720  
GCTCATGAT GAGTGTGGG GGTGTGTGG TCACGACCAC TCCAGACCA ACATGTCCAC 780  
AGCTTCCGTG GACCCCATTA ATTTCATCA CAAGATCAGA AAGGCTGGG TCATCAGCAT 840  
CTCCGAGCC ATGACCTTCA CGAGCAATTA GTCCATGGAG ATCGAGTGT TGTGTGAGCC 900  
CGACCTGTT GTGGACACT CTCAGAGCG CTACCGGGCC GCGAGTGGT TCTTCACTTA 960  
GTGTGCGTG AGCCAGAGG GCAAGTGGT GCTGTGCCC CAGCTGGTGC CCGAGAGCCA 1020  
GACGAGAG AGGCGTTTG AGGAGGCCA AGGGGGTAC CTCAGATGCA AGCGAGGCR 1080  
ACAGGGCCAC GCGGAGTTC AGCCCTAGAC TCCCTGCTCC TCCCACTGGT GCTCGAGTA 1140  
GCCATGCCA CGGGCCGAGT GTCCAGTAC TTAGAGTTC CCGCTTTGGC CAANACCCA 1200  
ATTCAATGT AGAGTGGTG TTGTCTGAAG TTTTGTATC ACHGTGTTAA CTTGTACTCT 1260  
CTCTGCCAA CTTACAGACC AAGCTTTT TTATGTAT TTATGTAT CAGTATGAA TGTACACAG 1320  
TGTGTGCG AGCGCCGGA GCGTTGGGC AGAAGCGTC GCGAATGCTT CCGAGCAGCC 1380  
TTAGGGTAT GCGAGAAC CAGCAGACT AATTAAGCTG CTGCTTGGCT GCAAAAAA 1440  
AAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1500  
AGAAAAAA 1509

35

(2) INFORMATION FOR SEQ ID NO: 180:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1316 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

AGCTGTATCA TAGGAAGAT GGCACAGG CGGTACACG TAGTGTCTC TCCGGCCAG 60  
CGAAGCCAG TCCCGGGGCG GCGCCGAGCG TCAGTTCCAG GCGCAAGCC AGCAAGCGCT 120  
GCGGCTCGG TTCCGCTGCG GCGTCAAGCC TGCATCTTCA GACCTTGGCG CAGCAAGGCG 180  
TGCAGTGGG GCTGCTGGC AGAGCCGGC CTCAGGCCA ATCCAGCCA GACCCGAGG 240  
CGCCCTCTG CTGTGCTGTC TCTTCAAGG CCGTTCCCG GCGCCGCGT GGTCAAGGTG 300  
CAGCCAGTCA TTTTGGCTTC CATTTGTGAC AGCTAGAGA GACGCAAGCA GGTGCTGCC 360  
CGAGTTATCG GCAAGCTGTT GCGAAGTCTC GACAAAGACT CAGTGGAGT CACCAATTC 420

60

TTTCAAGTCC CCGCAATGCA GTCAAGAT GAAGTGGCTG TTGACATGCA ATTTCCTAAG 480  
AATATGTATG AACTGCATTA AAGGTTTCT CCAATGAGC TCATCTCTGG CTGTACGCT 540  
ACGGGCCATG ACATCAGCA GCACTCTGTG CTGATTCAT GAGTACTACA GCGCAGAGGC 600  
CCCAACCCC ATCCACCTCA CTGTGACAC AAGTCTCCAG AACGGCCGCA TCGACATCAA 660  
AGCTACGTC AGCACTTAA TGGAGTCCC TGGAGGACC ATGGGAGTGA TTTTCAAGCC 720  
TCTGACAGTG AATACCGCT ACTAGACAC TGAAGCAATC GAGTTGAGC TCATCATGAA 780  
GACCTGCTTT AGCCCAACA GAGTGTGG ACTCTCAAGT GACTTCCAGC AAGTACGAGG 840  
GGCATCAGCT CCGATCCAG ATGCGCTGAG TACAGTGTG CAATATGCA AGATGTACT 900  
GTCTGGAG GTCTCAGCTG ACATACTGT GGGCGCTTC CTGATGAGC TGGTTAAGCA 960  
AGTACGAAA ATAGTTCCG ATGACTTTGA GACATCTC AACGCAACA TCATCAGCT 1020  
TTTCATGCTG ACCTACCTGG CCAAGCTCAC AAGTCAAG ATTGCACCTA ATCAAAACT 1080  
TGTAAACTG TGAATGAGC CCAAGCACTA CACTTCTCTG TCTAGGTATT AACCCAGGA 1140  
CTCAGAGTG AAGGAGAAAT GGGTTTTTGG TGGTCTTGA TCACACTGAG ATAGTCAATT 1200  
GTGTGACT CTAAATAAG GAGCTACCT TTTGTAAATT AAAAAAAAAA AAAAAAACC 1260  
SERGGGGGG CCGGTGCCA TTSSCCCTTT NTAAATCTG TTAAATATCC CCGGCC 1316

(2) INFORMATION FOR SEQ ID NO: 181:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 777 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

GCGATGKICA GACATGACTT CTATGTCGAG CTGTGTCAAG TGGCAGGGTC ATGAGGCGA 60  
CATCGATGAG GGTGCTGCTT ATGCTCCGTG CTCTGGAAAT CACCAAGGGG CTATCTCGT 120  
TTATGGGCT GGGCACTAGA ATTGCATGCT TCAAAACCAT CAGCTCTTGG CCAACAAGTT 180  
TGACCCAAAG GTAGATGATA ATGCTTTCA GTGCTTAGAA GAATACCTAC GTTATAGGG 240  
CGATTCTATT GGGACTGAA CTTTGAAGAC CACATATTG AAGAGGCGTT GCTTACCTGT 300  
TGGGGGCCAA GAGCGATGTT ACCAAACATG GTTCAGAAAT TTGGYKGGG AICARIBKGG 360  
GKGGGARRH CHRGGYTTG SCAMTTCSK KGGGWCCTT TTAGGCTAAR RRGCGCKGTW 420  
ATTAGATTCT GGTAAAGTA GATCTTTTG CCGTTGCAA TTTGCTGCTT GGTGTAATGY 480  
TGTGTGTGTT TTTGTACCC CTAAAGCTAG TAGTTCTGTC ACTAATCTTC TCATTAAGTG 540

AAATGAGAA CTGCTGTGAT AAGAGAGACT ATGTGTGAGA GCACTGTATT 600  
TCACTGTGAT GCTGCTGTGT AGCTTTTCCA TTCTGTGAG CTGCGCTTCC TAAATATCTCC 660  
AGCTTTGTGA GCTGTAGAGA GAATTTGTAT GAAATAGAGA CCTTCCCTTC TGTACTGTGA 720  
ACTTAAATAT AATAGCTTCC TGAATCAAG TAAATATATA AAAAAAAAA AAAAAA 777

## (2) INFORMATION FOR SEQ ID NO: 182:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 791 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

GGACAGATA ACTATGTAGA TGTATCTCTT AATATTTT TTAATTTT TATCTTGGC 60  
ACTGTCTTTC AATGTGTAC ATGTGTGCA GGAAGAAAT GCTTCTGT TTGAAATTT 120  
GATCTTTAG ACTGTCTT TTTCCTATCT TCTAGATCCC TGGCTTCTCT TGAAGGTACT 180  
TCCCTGAGCA GAACCCCTCC AGGTCAAGAG CAGAAAGAA GCTTCATGAG TCAACAGAC 240  
AAATGTGGCC TGAATATCTA TGTATTTGAT TTTCCTTGA ATTTCTGTA TGTTTTACT 300  
GATCTGTGGA AAGAAATATC TTGCAGAAA AATATATAG CTTTGAAT GTTGAATAT 360  
CTGTACAGC CTTCAGAGAC ATTCTTATTA TGTTTTCTCA GCAAAATGAG TTATATATCA 420  
TTTGTGCTCT GAGTATAGAG TCTCTAGAAA TCAAGAACT TCACTGATTA AATCTTACT 480  
TGGCTCTGGA CTGAGCTCTG TGAAGTGGC CTTCCTGACT GTCAAGATA GGTGTAGACA 540  
GACGACACT ACAAGAGCA CCTGTGTGGC TGTGTACTAG GAGCGAACT CCTGTAGAGC 600  
GCAATATAG AATTCAGCTT TTACATGAG CTGTTCATTA CTGTCTGAT GAACTGATTT 660  
TGTGTCTTCT CCTGTAGAC CCTGTCTCAT TTCAATGTTT TGTTTTATTT TTCTGTGTTT 720  
TTATATATTA CTAGAAATTA AATCTTAG GCTATCAC TTACTTATAT AAAAAAAAA 780  
AAAAAAAACT C 791

## (2) INFORMATION FOR SEQ ID NO: 183:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1405 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

AAATGATTA AGACTTGA AAGAGCTCT GATTGTAG GCTAAATAG CAGACTTCC 60  
AGTACATAG AACTGTAGAG ACTTGCACT GAGAAAGAG TGAAGAGGA GAAATATGAG 120  
AAGCTATAG GCAATATCA TCACTGAGA TCGAATTTAC AGATTTTGA GCAACAGAA 180  
GTATATAG CAGATCAC AAGAAACAG TTACAGATTC TGCATAGCA AATAGCTGG 240  
CAGAAACAT CCAACAGAG ACTTATAGCA GAATGAGC GACTTATCA GAAATCTCAC 300  
TATATAGAG AAGATCTTGA TGCACAAAG AACATTTCC AAGCAATAT TAAATATGTA 360  
GACAGAGAA TTCAAAATCT CAGATATGAG CTTACCAATTA AATCTTAG CAAATCAAT 420  
CAATCTGAT TGAATATG ACTTCATGAG CTACAGAGA CTCTCATCA GAACAGAC 480  
ATCTGTGAGA GTTCAGACAC AAAAAAGAC TCCCTGTCT TGCATGAGA GCGCTTCAA 540  
CAGAGATGA ACTCCGCTCT TGAATGAT AGTAAATGAT CTTCATTTA TATGTCTGTA 600  
ATTACAAAT GTTAAAGCAC TGTGTGCA AATGTCTCT TTCTTTTAA TGAACAGAA 660  
ACTATATGAG CAGATATGA CCAAAATTT CCAAACTAG CTATCTCAAT TGAATCAATTT 720  
AGTATTCGC TGGAAATTTT TCTCGAGA TACCCATAG CCGAGTTTT TGTATATTA 780  
TATATGCTT TCTTCACT CTGAGTATG ATTTCTCT TCACTTAC ACCGAAATG 840  
CAGAGACAC AACATATG CAAATGAC AACCCATTT GTTCATGTA TGTGTCTCT 900  
TTTCTTAC TGTATATG CAGAAAGCC AATGCTTAA AATTTCTAG AACATGAC 960  
AAGATATTT TATCATGAGA ACTTTTAC TTTTTATTT ATTACAGAG TATTTTACT 1020  
AAATCTTCA ATTCTTTA AATGTAGA GTTCTTAAA CAGACATTA TTATACAGC 1080  
TCACTTAC TTATCTAG TTATATGTC CTAAATATTA TGTACAGAA GATGTGTGG 1140  
TTGTGACT CTATACAG CACTGTATG TTATGTACA TGTATTTAG GTTATATCA 1200  
GGAATCTTA TTGTATAG TACAAATAT AAAAAATTT TTTTATTA TTCACTTCC 1260  
TTTGAATTT TCAATATTT AACTTTGCA AAGATTTAG TTGTATAG TTACAGCTT 1320  
GATGTGTGA AATCTTTTA TAAATATTA AATTAAGTA AATATCAT TTTCCTTTC 1380  
TAAAAAAAAA CTGCA 1405

## (2) INFORMATION FOR SEQ ID NO: 184:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1596 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

413

414

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

5 GTATCATGAT GCGCGGAGAA ACTGTCTCT TTGAGGCTGA CCTTAGGGC CCGAGAGGA  
AACTGGAGG CGAGGTGAC GCGGAGCGA GCAATTGAGA TCTCTCTGGT AGACTGGTG  
CACCACACC ATGTGGCTG CAAGCTGGT GTGTCTCGG ACATCACTT CTAGGGTTT  
10 CACCCAGCT TTACACAGG CTTCCCTCT TGTGAAGAT TCCATCAGA AGAATCAATG  
GCTGTACAA CTTAGCAGG AATTGCCAC CAAACAGAA ATGGGATCC GCGCTGGAG  
15 AACTGGCAA GAATCAAG AGCCAGAT GAAACATCG ATGGAAGAA TATTTAAT  
TGATCAGAT GGAATATGT TTTTGTCTG AGGGCTGCT GTTGTCTTG GAGCAATGTG  
420 CTAATATGC TTGGACTGT CTAATGAGAT TCGACTATT GAAAGGCTG TAATTTGGCC  
TGATATATC AAGATAGAA TTCAATCCAC CTAATATGAC TTAGCAGGA GTATTGGTTT  
480 AAGAGTTTG TCTCCATAG CAATCAGAG AAGCCCTGT CTCATGAAT TCAATGATG  
540 AGGCTCTTG GTGCAATG GTTGACTTT TCGAGCCATG GTTGAGCTG GAATCTGGT  
600 ACGATCATA CCATATGACC AGAGCCAGG CCCAAGCAT CTTCCTGGT TCGTACATC  
720 TGGTGTATG GTTCAGTGG TGGCTCTCT GACAAATTA GGGGTCTTC TTCTCATAG  
780 AGCTGCATG TACAGAGTG GATTTGGG AGGCTCTCC ACTGTGGCA TGTGTGGCC  
840 CAATGAAAG TTTCAGACA TGGGTGACC CTTGGAGTG GGCCTGGCT TGTCTTTGT  
900 GTCTCATTG GATCTATGT TTCTTCACC TACCAACCGT GCTGTGACA CTCCTTACT  
960 AGTGCATAG TACGTTGAT TAGTCTCTT CAGCATGTC CTTCGTATG ATACCGAA  
1020 AGTAATCAG GTCGAGAG TATCAGCAT GTATGAGTT CAAGAAATG ATCCCATTA  
1080 CTCGATCTG AATATCTACA TCGATCATT AATATATTT ATCGAGTGG CAATATGCT  
1140 GGCACCTGA GCGACAGAA AGAATGAG TCACTCAGT TCTGGCTCT CTCCTACATC  
1200 AATATCTTG TTATATGGG CAGATATGA TTAATATGT TGTACAGCA GCTTCTGTG  
1260 AAGTTTAAA GATAAGAAC ATGTATCAT ATTTAAATG TCGGTAAATG TGATGCTCA  
1320 GGTCTGGCTT TTTTCTGGA GAATAATCC AGTAATCTC TCCAAATTA GCAACACAT  
1380 TTTCATCTC CATCTTGAG TGAATTAAA ATGTTTGGT GAATGTGAA ACTAAAGTTT  
1440 GTGTATCAG AATGTAGTC TTTTCTTAC TTTAAATTT AGTAGTTCA CTCGATTAAT  
1500 AAAATTTAG AACCTGTGT TTGCATAT TTGKGAGTG CAGATATG TAATTAAGC  
1560 ATTCCAGTAA NAGTGTNTT AAGTGTGTC TATATN  
1596

60

(2) INFORMATION FOR SEQ ID NO: 185:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2293 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

5 GCGAGAGCC GTACGAGCA GAGCAGGAC GACAAGGCG ACTCGAGGA AAGCGGCTG 60  
ACCTGATGG AGAAGTGTCT CTTCTGGGC CTCAGAGCC GCGAGGTGA CACATCATTT 120  
TGGATGACT GTATATCATC TCGATTACCT GCGTGTATGT TATTGTATTT AGCATTTGGA 180  
GGAAGTTAC AACTAGAGCC TTGTGGAATG AGAGTAATA GTCTATTAAAC AAGAAGGTA 240  
ATCTGTAGT CAGATGCTCC AACHGGGAT GTTCTTCTTG ATGAGCTCT GAAGCATGTT 300  
AAGGAATCT AGCCTCCAGA AAGGTCCAG AACTGGATG AATTACTTNG TCGTGAACA 360  
TGGATTCAT TAAATTTGCA TTATCAGTAA AGAATATPAC GCGAAGCATT ACCTAAAGAC 420  
CTGTGGAHA AGGTTGTATT GACACAGAG AATCAGAACT TCTTACTTT TGACATGACA 480  
ACATATCCC TCACCAATTA CAAATTTAG CAGGCTCCA TGAAGAAAT AGAGGAGGCC 540  
GTCTTTGACA AATGGGTGAA TGACCTCAC CCGATGACA GCGCTTCTC GCGCTCAT 600  
TACTGTGCTC ATGCTCGGA CTTCTGGAG AATGCTTTTG CTCCTCTTCT GCGAGAGCAG 660  
TATGATTTGG CTACCAAGAG AATGGGCGAG CTTCCTGACT TAGACCTGTA AGTGGATGTT 720  
CTGAAGGCC ACACAATGA GGTCTGTGG GCGGTGTGG CCGGTTCAC CAGTAACTC 780  
TCTTGGGGT GAACATCTC CTTTCTCTC AAGTAAGCA GTAGTTTTC TTCTGTGAC 840  
TTCTGGTTT CTGTATTTTG TACTTCCA CACTATATTT GGCCTCTGTT TTACAAATG 900  
GTGGTGGCT TTTCTTTTTT TGTAGTGTG CAGGATCTG CTGTAGAGAG AGGCTTCTC 960  
CTTCTGTTT TTAAGAAAG TTTTACTGCC ATATGGCAT TCCATTCCT GTTGCATCC 1020  
TGACTGTAC CTGTTTGGG TTTCTGCTC ACTTGGACT TCAAGTAGC TCCAGGCTCC 1080  
TCATACGAC AGCTTTTGA TGACTTCAG TTGAGTTCT CCAATGTGC ATGTACTCT 1140  
AGCATTTGC CTACAGTTCA GACAGATG ACAAAAGGC CTTCAACTCA CCAAGGTAA 1200  
ATATCTGAT CTATTAGAC ATTTTITACA TAGACTTCAG TTGAGTGTG TACTTAGCAA 1260  
AATTATTTT AATGTGAAC AGCAGATTA ATACTTAATA TAAATGTCC CTGTGATTTT 1320  
GCTTCCCAT TAAATCTATT GTATTATAC ACTGTATATA ATTTTACTA TAAAGTCCA 1380  
ATGTTTTCAC AGAGCCAGTT TGGATGGGC TGCATTCAT TTATCTGTA TATAGTTGA 1440  
60 ATATATATA AATTACCCCT TCTTGTGGC ACCCTGCTC CCATCTTACT ATTTTGCAG 1500

ATCTAATCAG TTGTACACCT GGTGCCCCCT GCTTCTCTCA ATCAATGTTA TTGTATGCA 1560  
AAATGACCT CTGTGCTCTG AAGAGAGAG AAAAGATGTG TGTCTGATG GTTCTGGAAT 1620  
TTTTGACCT GTGCATTTA TGTACTCTT TGTCTATGCA TCCCCCTTTT AGATTTTTT 1680  
TAAATTTAT CTACTGTTT TTAATTTTC TATGTGAGAG AGGCTTTGCA CGATACCAA 1740  
TCTGTAGTT CTTTTCTGT CCGAGGTAA ATTAATATCT GCTGTGAAT GTCAATTAAC 1800  
TACTGACCA TTCTTGAGCA AAAAATCAA ATGTCAATCT TACAGATGT TGCATGTAA 1860  
TTGTATGCA GTATATGTTA CAGCCAGAG GATTAGAAAT TTGTATACAG AAAGCTAAT 1920  
GTTTAAATTT TTATATACA ATTAGATTA TTACATAT CTAGACTTAA AACTTTGCTT 1980  
TTTAAAGTT ATTTTACTA TTCTTTTATC ACTTTATGT ATCAACACA TTGATTTGAT 2040  
AATGTAAATA GTATATGTG AACAAATTA ATGTCAAAAT TTTTATATTC CAGATGTCAAT 2100  
GTAAATAGTG GGGCTTCAAG GTTTTGAAG ATTTTTTGT TGTATTGTA CATTCAATTC 2160  
AAAAATCTA GATGTGTAT AACTGTAGAG TTGAATTTTA AGGATTTCC TAAATGTAT 2220  
ACTATCTTT TATCTAGAT AATAAATAA CAGATATCT GAAATGCTT AAAAATAAA 2280  
AAAAAAAA AAA 2293

(2) INFORMATION FOR SEQ ID NO: 186:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1212 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

GAGCAGAGGC GAGCCGAGGC ACCGTACCT GAGACATGTC GTTTCACATC GTGCCCTCC 60  
CCCTGGGTTT GCGACCTTTA GCGACTATTC CCGCTGCCC AACCCGAGCTG CGAGATGAG 120  
GCGGTGCTG CTGCTCCCG GTGATGTAG GCGGTCTCTT GGGCTTCAGG CTGTGTCGAG 180  
GTATAGGTG GATTCACGA CGTTTCAGG CCGAGATCT CCGCAAGT GGCAGCATG 240  
GCGCAGAG GAGACATG CTACAGGAT GTCTGAGTG GACTTGAGAA TGAAGAAAT 300  
GAGATAGAG AAGACAGAA TTATTTAGAT GGCCTGTG ATATCTGAT CTGATTTGAA 360  
CAATTTGCT GAGGTCTTA ACGTATTA AAAACAACA GATGATTC TGAACAGAT 420  
GTCAATGAG TGTAGAGAG ATCAAGGAA GTCTGACCTT ACCAAGCAG TGAAGAGAG 480  
TACATTTACA TTGCAGAG GAGATTTA ATTTCTGTCC TTGTATACAG AGCAGATCC 540

TGCAGGCTC TGTCACTTG CAGGAAATTC AGTTCAATGA GATAGAAAT TTCTTACAA 600  
ATACATCCC GAGTTTATGA AACATTTCA TTATAGATA ATTAATGTGA GCACGTTTAA 660  
AAGATGTG AAGCTGTGT ATTCAGAGA ATATGAATTT GCACAGAA AGGCTGTTC 720  
TGTATGACA CTGTATGCA TTATGTAAAG CATACAGAG CTTCAGTTT ACCGAATTA 780  
CATCTTCAAG AAAAATATAG ATGAAAGAA GAGGAAATTT ATGAAATAG GCGAAATGA 840  
GAGACCTTG AGTTGATGC AGTTATGAG CTGCACATC ATGTATAT GAGGCAAT 900  
TGTGTGTCT TTTTCTCTC ACGTATAG CTTCAGAG GACCTGAGT TTAATTCAT 960  
CTCCAGATG ATTTCTCAG CAGACAGAC ACAGAAATCT ATTTTCTCC TAAATGCTG 1020  
TTTCAATAT GACAGAGAG CTCTTTTGA AGTACAGAT CATGTCAATC CTTGTATCA 1080  
TATATCATT TCTTTTAAA CATTCTTT TGTTAATA AATAATAG TAAATTAAGC 1140  
TATCTTATT GAAATGAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1200  
AAAAAAAA AN 1212

(2) INFORMATION FOR SEQ ID NO: 187:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1605 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

GTTCGAGAA GTTCTTTTG TCCAAACATC GGGGTTCCT CTTTTGTGT TCCGCGCAT 60  
GCACTCTTC CTGCACCTG GAGCTCTTAC TTCCGAGAG CCAATCTTG CCGACTCCG 120  
GCGCGGAGCT AGCGCGGTT TCAAGCAGAG GAGCTCTTAA GGAACATGC AACTACAGAG 180  
GCGCGGAGCT GCGGCGCTCG AATGTAGCT GGCCTGGAAT GCGAGGTTT CGAAGAAAC 240  
ATCCAGGCG GAGCTCACAC TGTGATTCAC ATGAGAGACA TGAATGATAC CTGAGGCTCT 300  
AGCTTGAG ATATGATGTA GCTGATCAG CAGCTGCGCG AGAAGAAAT AGACCTGAT 360  
GAGATGAT GAGTCTCTC TGAAGAGAG GATGAGAAAT TCTTGAGCAT GAAAGCTTT 420  
AAGAGAGAG TGAACGAGCA GTTGGCAAT CAGATGTGAG AGGCTGAGAA AACACAGCC 480  
TGCAGGCTCT TCACTTTGA CCGCAACAT GACATCTTCA GAGCTTACTT TGAATGAG 540  
CGTCTCAG TGCAGAGAG GCTCTCAG TGCATGATCC GTATGAAAT GTTCAGCTTC 600  
CCCAAGAAA TTGCAGTGA ACTGTATGA CCTGTATGAG TGGTCTTAC TGTGTGTCT 660  
ATCTTACTCT ATGTATGTA GAGCTTCACT ACTATTAATC GAGAGGAGC CCGATGAGCC 720

5 ACGGCCATG GCACTGCTT CGGCTACTG CTGGAGTCT CATCTCTCAT TTACTTCTT 780  
GCCTACTGT GCAGGCCA GATCAGCATG CTGCACTGT TGGCACTGCT GGGCTATGCG 840  
CTCTTTGGGC ATTGCATTGT CTTGTTCATC ACTTATATA TCCAGCTCCA CCGCTCTTC 900  
TTACTCTTCT GGGCTTTGCT GGGTGGACTG TCCACACTGC GCATGGTAGC AGTGTGGTG 960  
TCTCGAGCG TGGGCCCCAC ACAGCGGCTG CTCTCTGTG GCAGCCCTGCG TCCCTTACAC 1020  
ATGCTCTTCC TGTCTTATCT GCAATTTGCG TACCACAAAG TGGTAGAGGG GATCTTGAC 1080  
ACACTGGAG GCGCCACAT CCGGCCCATC CAGAGGGTCC CCAGAGCAT CCGTGGCATG 1140  
CTCCCTGCTG CTGGCTTCC CAGCACTGCT CTCACGCCA CAGCCAAAG TGTTCGGTG 1200  
ACCTCCAGT CACACTGACC CCACTCGAAA TTCTTGGCCA GTCTCTTTC CCGCACTGCG 1260  
AGAGAGAGG AAGACTATTA AAGCACTGTC CTGATGATAT GTTGTGTAGA TGGGGTTTGC 1320  
AGCTGCCACT GAGCTGTAGC TGGTAGTA CTCTCTTGAT GCTGTGCGC ACTTCTGAAA 1380  
GCCACAGGC CAGCACTCC TGGCCAGAC TCCAGGCTC TCCAGCCAT CAGAAAATG 1440  
GGTCAGTCC TTTCAGACC CTTCCGACC TACCCCTTCC TTCTCTTTA TCTCTGCCAC 1500  
ATTGTCTGC TAATATAGA CTGTGTATTT AATATGTGA TTGAAGTCTG GAAAAAAA 1560  
30 AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAAA TCGAG 1605

35 (2) INFORMATION FOR SEQ ID NO: 188:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1516 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

45 ATTTCGCATG AGGGGTGAC GTGTGGCTG GCGCGGGAA ATGGGGGCTT CAGGAGAG 60  
CGGACTTCA GCGCGGGAG GCAGCCGGA GGAAGCATTT ATGACTTCT ACAGTAGGT 120  
GAACAAATA GAGAGAGAG ACTGGTTCTT AACTTCGAA ATACGATTT AAGACTGAC 180  
CGGTCTGCT TCTCTTACT TCAATTTGA CCAATTTGAG GTTCTTGA TAGATCTGA 240  
AGTTACAGT GAAGAAATA AAGAGAGTT TGGCAGTTA TCAATCTG TCACTCTGA 300  
55 CAAAAATCA GATGATCTG ACAGAGACA AAGGCTTTT GAGCTGTG ACNAGCTTA 360  
CAGTTCTTA CTGATCAGG AGCAAGAA GAGGGCCCTG GATGTATTC AGCCAGAAA 420  
AGATAGCTG GACACACTG TGAAGGCG AAAAAACA TTAAAGAGC AAGGAAAC 480

TACATTTGA GAGGAGATG ATCTGAGCT GTTCMAACA GCTGTATATA AACAGACAT 540  
GAATCTTTT GCAGAGCTG AATTTAAAG GAAGAGAGA GAAGCCAAAG AGATCATGA 600  
5 AAGGAACGA CAAAGGGAG AAGATTTGA AGCTCAGAA AAGCCAAAC GGGAAAGGA 660  
GTGGCAGAA AACTTTGAG AAGTGAGA TGGTGTGTG GACAGTGGC GAACTTCCA 720  
AGCCATPAG AAGGGAGA AAGAGAGAA AATTCGACC TTCTTGAGC CACCGAAGT 780  
AAAAATGAG CAACCTGAT GACCGCCAA GGTCAAGGC ACAGACCTT TCCCTCTTA 840  
TCTCCCTTC TGTTCGAG GACTCATCT TTCTGCCAC TTCCAGCCCA ACATAGATTA 900  
GTATTTGCT TTATGTCAT TTGTTTTCA ATAGATTTA ATATCGATCA GAGTAATCT 960  
TTTGTATTT GAATGAGGG GCTTGGTTA AAAAAAGAC TTTCCTCTTC CTTGCCCTA 1020  
GACACAGG TATTAGAGG TCCACCATT GGTGCTGCT TCTCTTCCA CAGCTGTAA 1080  
CTCAGTCTT TGTACTTAC TGAATGTGA TGGTAGAAA CTTCGTGAT AGTTTGTGA 1140  
AATATCCA TTAAATATC TGTTAATAC AGTGTGCTG TCACTTCCA GACAAGCTG 1200  
25 GAAGGGCAC CTTAGAGGC CCTTGGCTT CAGTTGCTG CTTCGGGTG TCTTCCCTTC 1260  
GAAGGCCAG ATAGACAGG GAACACTTGT GACACACAG AGCAGCATCT GATGCCCTT 1320  
GGTGTGGC ATGTGCCCC TGTACTTGA CCAATCAGT TGGCATGAG CCCAGGCCAC 1380  
CCAACTTTT CACTTCCA AGAGTAGCC GTCTCCACC CAGTACCATG TCTTAGCTG 1440  
TCTCATTTG TTAGTGTAA TATCTTTAT GTATATATA TTTTATACC CAAAAAAA 1500  
35 AAAAAAAAA ACTCGA 1516

40 (2) INFORMATION FOR SEQ ID NO: 189:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 681 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

50 GCTCCCATG TCTGGCTGT CCGTACATCA CCGTGTCCC TCCAGAGGG GCGTACAGC 60  
CATCTCCCT CTGTAAGGCT CTGACTCCC TCCACTTTG GGGCTCAGC TTATCTGGG 120  
CAGGGGACA TTGCAGCAT CTCCCTGCT CNGGACTCA GGTGCTGAG TATAAGGCTT 180  
GGGCCGAGA TCCCTGTGA CACTCTCTG GAGAGACTC TCANAAGTGA CTGTATTTT 240  
GAGTTACCA GCATTAATC CCACTCTG AAGCAGGCTC AAACCCMAG ATCCAGGCT 300  
60 CCGTGGGCTC TGTGGACTG TCTTCCGAG ATCTTCTCTG TTGCACATG GGAACCTTA 360

GAAGAAAAAG AAGAGGAGCCT GCTTGGCCAG CAGTGCAGAG CAGTCCAGC 420  
TCTGTCTGCC TGGCTGGAAAT GTGGGAGCCT GCTCCGCGTC AGCTTGTCT GTCTGTACAC 480  
TATGTTTACA TCCCGAGAGG GTTTCCTGCT CATTCCAGAG CAGTGCAGAG TGTGTCCAG 540  
CAGCTTCAGT TGGGTGTCTG ACATGTGTCA CAGTCCAGC CTTGAGCCGC GGGGGGTGTCA 600  
TGTGTCTCTC CTGAGTCTGT GTCTGGAAAT TTTTCTTTT GAAAAAGCTT TTAATATTAA 660  
AAAAAATAA AAAAAATCTG A 681

15 (2) INFORMATION FOR SEQ ID NO: 190:

20 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1014 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

GGCTGAGCC AAGCAATGGA TAAATTTCTG GAACTGTAA ATTGAGTGT TGTACAGCA 60  
GTATGTTTAC AAACATTTGC AACAGACACA TCAAAACAGA GTCTGCCACT TTTGTCTATC 120  
GCTGAGTAAT TGTCTGGACA AGTGTGCTT TTAAGCTCT TAAAAAGCCA GAAAAAGCT 180  
ATTGCTTAAAG CAGACATGAA CAGATTTGAT TAAATGTCAA AAGCCATTTG CTTGATGATG 240  
TAATGGCTAC ACTTCTGAC AACCTCAGAG TTCAAGAGAG GCAAGACAGA GTTGTACACA 300  
CTGTAGCAAT AGCTATGTT GCAAGAAACAT GTTCCAGCTT TACAGTACT CTTGCTTTAA 360  
TGAATGAATA CAGATTTCT GAACTGAATG TTCAAAATGG AGTGTAAAA TGGCTTTCT 420  
TCTTGTTTGA AATATTTGAT GAAATGGGAA AAGAGTACAT TTATCCCTTA ACACCTTTAC 480  
TTGAAGATGC TTTAATGATAT AAGAGCTTGG TAAACAGACA GAGCCATGAT GCAATGTAC 540  
AAGACATGTC ACTTGGGCTT TATGGAATTTG GTTGTGAAGA TTGGCTGAAT CAGTGTGTA 600  
ACTATGTATG GCGCAATGTA TTGAGAGAT CTTCTCATAT AATTCAGACA GTTAAGGAGAG 660  
CCCTAGAGAG CCTAGAGATT GCTAATGCAG CAGTGAATAT GTTCAAAAT TGTTTACAG 720  
GTCTGTTTCA CCGAGCCCGG AAGTGCAGAG ATGTATATGG GAAATTTTAC AATCTCATCT 780  
ACATTTGTTG CAGAGAGCT CTTCAATACAG ATTACCCAGAG AATTCAGACA CGAATGATAG 840  
TAATGCAGAG TACTTGCAG CTTAAACTTG CTTTGAATTTG GTATGTATAA CTTTAAACA 960  
TTGCAGATCA GTTGAAGACT GGTCCATAGAG GAAAGAGTAA GAAATTTCAAT AAGC 1014

5 (2) INFORMATION FOR SEQ ID NO: 191:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2779 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

TGGCAGAGAG GTGTGTCCAG ATGTGTATGC TGTGTGTCT AGCTTGTCT GACAGGGAG 60  
AGTTAAAGTC CCGTGTCTCA CCGTCCCGAG TGGCCAGATG GCTGTAGAGGT GACCCAGAGA 120  
CGAAGACTTG CTTGCTGAGG CTTAGTGTCT AAGACTTGGG AAGGAGGTTT CCGCCATCTT 180  
CTGCTGTCAA CCGCGGAGAG CCGTCCCGAG TTTCATTTCT CCGCAGAGCC CTGTATCTGA 240  
GAACTTCAG GATCCAGACA GCGCCAGAG CTTGAGCTTA GCTTGGAGGG CTTCAGTCCAG 300  
GCCAGACAG AAGCCGATAG GAGAGAGAC AAGAGCTTGG ACATCTCAT CTGCAGAGAG 360  
GTCTGTGATG GACGAGAGAG CCGTCTCTCT GTAGATAGAC CTGACCTCTC AGCTTGTCCAG 420  
TTTTCAGATT GAGAGCATTA GATGGAGAGC AAGAGATAG CTGTAGAGGG GACAGAGGAA 480  
AGCTGAATTT TGGAGAGAGG CTGCTCTCCA TGGATGTACA GTTCCAGAGC GAGAGACCGA 540  
AATTCGAGCC TCGATGATGA GTTCACTTA ACTTACGAAA GGAACAGAT GCGATGTACAG 600  
CGAATCCAAA CCGATTTTAC GAGATTTGAG TCTTCAATGC GATTCCTCGG GATTCCTCGG 660  
AAGATCTGAG TGGATCTTCA GATGATCTGA GCAAGACAG CAGTACTTC ACCGACTGAG 720  
AATACAGACA GGCCTCCACA GATTAAGAT GCTGTATGAA GCTGTGTCTG AACCTTAAAG 780  
ACGGATCTGA TGCCTGATTT CTGCTCACTG TGCAGATCTA CCGGAGCTT GCGAATCTTC 840  
AAGCCCTGAT AATATCCAG TGCAGATAG ACTATTTACAG AAGCCAGAGC GCTGTGCACA 900  
TCCCATTTGA GAAAGAGAGT CTGAGATGAG TGAAGCTCT GATGAGAAAT GGGGGCAATG 960  
TGCATCCAG GATTCCTGAG GATTCCTTCA GAAAGGCCAA GGAATTTGCT TTTATTTGAG 1020  
TGAAGTACCC CTCTTTTGG CCGCTTGCAG CAGAGATGAG GATGTGTAA GCTACTCTCT 1080  
GAGAAACCAA CAGAGAGCC CAGAGCTTCA GGCATGACT CCGAGGGCAA CAGATGTCTG 1140  
CATGCTTAG TGAATATCT GAGACACTTA GCTGAGACA TTGCATGCT GAGCAGCATG 1200  
TATGATGAGC TGTCTCAAGC TGGGGCCGCG CTCTCCCTTA CCGTGCAGCT TGAAGACATC 1260  
CGAAGCTGAG AAGATCTGAG GCTGTGAGG GTGGCCGCA AAGAGGGCAA GATTCAGATTT 1320  
TTCAAGCACA TCTTCAGAGG GAGATTTTCA GAGCTGAGCT ACTTTCAGG AAGATTCAGC 1380  
GAGTGTGCT ATGGAGCTGT CCGGATGTG CTGTATGAGC TGGCTTCTGT GACAGAGCTT 1440

5 GAGGAGAACT CAGTGTCTGA GATCAATGCC TTTCATTTGA AGAGCCCGCA CCGACACGA 1500  
ATGTCTTTT TGGAGCCCT GAGAAACTG CTGACGGGA AATGGAACT GTCTATCCC 1560  
AAGTCTTCT TAACTTCTT GTGTATCTG ATCTACATGT TCTATCTAC CCGTGTGCC 1620  
TACCATGAC CTACCTGAA GAGCAGGCC GCGCTCAC TGAAGCGGA GGTTCGAAC 1680  
TCCATCTGC TGAAGGCGA CATCTTATC CTGCTAGGGG GAATCTACT CCGTGTGGC 1740  
CAGGTGTGT ACTTCTGGG GCGCAAGTG TTCACTTGA TCTGTTTAT AGACAGTAC 1800  
TTTGAATCC TCTTCTGTT CAGGCGCTG CTACAGTGG TGTCCAGGT GTGTGTTC 1860  
CTGGCAATG AGTGTACTT GCGCTTCTT GTGTCTGGC TGTGTCTGG CTGCTGAC 1920  
CTGCTTACT ATACAGTGG CTTCAGGAC ACAGGATCT ACAGTGTAT GATCCGAG 1980  
CGCTGTGAG CGTGAGCAG GANITGGG CCGCGAGCT CTACAGGCC CCAATGCCAC 2040  
AGATCTAGT CAGGCCATG AGGACAGGA KGAAGAGGCC AAGCGGGCC AGTACAGGG 2100  
TATCTTGA GCTCTCTTG AGCTCTTGA ATTCAACATC GGCATGGGG AGCTGGCTT 2160  
CGAGGACAG CTGCACTTCC GGGCATGCT GTGTCTGCTG CTGCTGGCT AGTGTCTGT 2220  
CAGCTATAT CTGCTGTCTA ACATGCTAT GCGCTCATG AGCGAGACG TCAACAGTGT 2280  
CGCCTATC AGCTGTGGA TCTGTAGCT GCGAAGGCC ATCTCTGTCT TCGAGATGGA 2340  
GAATGCTAT TGTGTGTGA GGAAGAGCA GCGGCGAGT GTGATCTGA CCGTGGCC 2400  
TAGGCAAT GCGAGCGCG ATGAGCGTG GTGCTTCAG GTGAGGAGG TGAATGGCC 2460  
TTCATGGAG CAGAGCTGC CTAGCTGTG TGAAGCCG TCAAGGGGAG GTGTCTCTG 2520  
AACTCTGAG AACCTGTGC TGGTTCGCC TCCAGAGAG GATGAGATG GTGCTCTGA 2580  
GGAATCTAT GTGCGCTTC AGCTCTCTCA GTCCAACTGA TGGCCAGAT CAGCAGGAG 2640  
GCCAGGAC AGAGCAGAG ATCTTTCGA CCACATCTGC TGGCTCTGG GTCCCAATGA 2700  
ATTCTGTGG CAATATATA TTTTCTATA CTCAAAAA AAAAAAAA AAAAAAAA 2760  
AAAAAAA AAAAAAGC 2779

(2) INFORMATION FOR SEQ ID NO: 192:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1923 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

5 ACCGCTCCG CTCGCTCG CTGCGCCG GCGCGCCGT CAACATGAT CCGTGGCG 60  
TGGCTGGA GCGTGGCC TGGATCTGC CCGTCTCTT ACTCAGGCC ATCGCTTGG 120  
ACATCATGC GCTGGCGCG CCGGCTGCT TCGAGTCTAG CGACCAAGG CAGAGTCTT 180  
CGCTGTGTG GAATGTCTC CAGAGGGGG CCGCAGCGG GTCTTAGGAG GAGGCTGTG 240  
AGAGCTCAT GAGTACGG TGGGTAGAG CAGCGCTGC CATCTCTTC TGTGCTTCA 300  
TCACTCTGT GATCTGTTC ATCTCTCTT TCTTGGCCCT CTGTGAGCC CAGATCTTG 360  
TCTTCTGAG AGTCAATGA GTCTCTCTG CCTTGGCTGC TGTGTTCAG ATCATCTGC 420  
TGTATTTTA CCGGTGAG TACACCGA CTTTCACTT TCAATGCCAC CSTGCTTCA 480  
CTTACATTA TAAGTGGCC TACGGCTTTG GTTGGCAGC CAGCATATC CTGATGGCT 540  
GTGCTCTCT CTCTGCTGC CTGCCACT ACGAGATGA CTTTCTGGC AATCCAGC 600  
CGAGTACTT CTACACACT GCTTACTTG GGAATGAATG TGGAGAAA TCGCTCTGC 660  
TGAATGAC TCGAAGAA GAAGTGTCT CTGAGCGA CTTTGAACC ATTTTGTGC 720  
AGTGTCTTA TTTTAACT AGTCAAAAT GTTAATATA TTTGGAGAA ATATTTTCT 780  
AAGTATGTT ATATTTCTT TATATCTTT TATATCTTT TGTGAGTGT TGTCTTTCA 840  
CTATTAAT ATACTATGC AATATTTCT TATATCTAT CATTAACAT ATACTACTT 900  
TGTAGAGA TATCGAGTG AACTTATCA CTTTATAGG TAAATATGAG GTTTCAGAA 960  
TTTATATAT TCAATAGTT CTGTATATT CCATATAGA TGGACTGCT CTGTATAGG 1020  
CTAGAGAA GAGGAAGTA AGTTTAAAG TGTATATCA CCATATATC TAAAGAAAT 1080  
CGAAAAAA AGTTTATTTT CAGCCTTGG AACTATTTAA GGAAGCAAA ATCATTTCT 1140  
AATGCATAT CATTTCTGAG AATTTCTAT TAATATCTG AATCATCAT TTCAGCTAG 1200  
CGTCAATTT GACTCGATAT GTCATCTAG AAGTACTAT TTCAATGGTC AAACCTGTG 1260  
CCATAGTGG TAGGCTTTC CTTTATGCT GAATATTTA GATGAATTT TCTCTTTAA 1320  
AGTCTTTAT AGGTATAGG TGTGGGAAA TCTATATTA ATAAATCTGT AGTGTTTGT 1380  
GTTATATGT TCGAACCG AGTAGACTG ATTGAAGAT GGAATGGTC TAATTTATCA 1440  
TGAATGAG ATCTGTTAA GTTGTATGT AAGCATTAG GAGGTCTAT CTGTGACAA 1500  
AAGTGCAT AATCAGCCT CAGGCAATA AATGACTTC TTTTCTAAT CTCAGTTTA 1560  
TGTGGCTCT ATCATATAGA CAGCCTTCT ATAGTTTGA ACTGTATGA GAAGCTTCA 1620  
TATAGTAAA ATCTGTGCT TTCTGTGAA ACAGATTTA AATGTCTGAT ATAAACATG 1680  
CCACAGGAG ATCGGGGAT TTGATTTCT CTGATAGCA TATATATGAT GCATCGATA 1740  
GTCATATG ATTTTATACC ATTTGACTT ACATATGAA AACCATTTCA TTTTAAATAT 1800



CGAGTATATTA TTTCGTAAAT TGTGAGAAAA GGTAAATTTTA GTTTTCAATTA TGAAGTTTC 1860  
CGAATTAACC AGGTATTTCTA AAAAAAAAAA AAAAAAAAACTN GAGGGGGGGC CCGTACCCA 1920  
ATT 1923

10 (2) INFORMATION FOR SEQ ID NO: 193:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2146 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

AGCGTCAGGG GAGACACTTC AAAATTACAC AGCTTTTAC AGCTGACAAA ATTGGGCTTC 60  
AGACCGAAT CTGGATTCAA GTCACTCATG GTTATTTGC GGCATTCCTT CCGGCATTC 120  
GACCTGACA TCTCTCTTC CGATGACAA TGAAGAGAC GGGGGCCGAG CAGCTGATG 180  
GCTGAGAGGG ATCAAGCTTT CTCTGGGGCT GGGCACTGAT AAGAGCAAT GCTCTGTGGA 240  
CGGCAATCCT GACTCTCAC CTGGCACTCT GACTGCGCTG CTAAACCGCT GTCTCTGTT 300  
CGCTTATGTC CCTTGACTAG CATGACCCGC CTGATGCAAT SCTGACAGC CTCTGTTTC 360  
AGCGTCATTC TGTGATATGC ACCGCAAGCC GCTTCCACAC CAGCGAGGCC ACCGCAAGCC 420  
GCTCCACAC CAGCGAGGCC ACCGCAAGCC GCTTCCACAC CAGCGAGGGG GCGCACTTC 480  
AGATCAAGCC CAGCTGAGCT GCTCTGTGGA ACGAGGCCGG GCGCAATTC ATGAGCTCT 540  
CGCTGTAGGC GCGTGTGACA GCTGCGCGCC CCGGACTGTC ATCTGTGAG TGGCGCCCTC 600  
CTCCCGGAGC CATTTCGCC CCGTGCCTGC CATCTGTGCG CTAACTCGAT ATTATTTCAA 660  
AGCTTATTTT GTAAAGATGA GCTCTGTGTC AGACAAATGA GATTATTTAC GTGGTGTCCC 720  
TCTGCAAGAG CCGGATGTGCG GTGAGACAAA GGAAGAGAGC AAGCATTCCT GATGTGATTC 780  
CTTTCATAT AACGATGTGC CCGTTGCAC TCTCTCTCCC TCCCTGAGAG ACACCAACT 840  
GCGAAGAAC AAGACGTCAG AGCAACAAT TCAAGAACCC AAGCTTAGCC GCGCTTAGC 900  
ATCTGTGCTT AAGCGGTGCG CTGTGTGAGAA GCGCAAGCCCG CAGTCTCCCG TTTCCTTTT 960  
AAGTGAAGAG AAGCTGATCC AGTTTTCGGA AACAAATCC TTTCCTCAT TGGGAGAGGG 1020  
GGTATATGTC AAGTGCAGGC ACCTCTTTTA AACAGCAA ACGAGAGGG GAAAAAGTC 1080  
GGAATCAAT CTAGGCTAGA GCAATTTTGA ACAAAGATTC TACGTATTTA ACTTGAAGAA 1140  
ACCGATTTT AAGTTGTGTC CATCTGAAA GCTTTGATG CAGAGCAA CAGCTTCAT 1200  
TTTCTAGCA TCTCTTAAT GTGAGAGAA AGCAGGAC AAAATCTCT GCTTTAGC 1260

ACAAAAATAT TTGACACAC GTTGCGATTC ATGATTTTC AAGCTTTTC TTCTCTTTC 1320  
TCTCTTCTT CCAAGAGCCC AACCTCCAC CCGTATACA TGAACAGTTC ATTATTTTC 1380  
TTGAGGAGA AATCATTTTA GATTGTTTT GATTTCTTGA GATGAGGG CAGATTTCA 1440  
CAGCTGCGCT GCTGTGATG ATGTCTTTC CAGGGCGCGG AATGAGACA CTGGGATGCG 1500  
GCGGAATTC GGTTCAGTC ATGTAAAGGA TTCTTTGTC TTGTGTGAG ATCCAGTCA 1560  
GTGTGATTT CTGTGATTC CAGCTGATT CGAGAAATT TGTGTGATG GCTTAATTC 1620  
AGTTTCAAT CTTCAGAC TCAGCTGAAA CTGTAACTCA GTACTGAC CTGTGTGAC 1680  
CGATCAAGTT CTGGAATCTT CAGAACTTT TCTCTTGTG GGGTGGGGGG TGGGACTCA 1740  
CTTGCGGAGC GATGCTGAG AAAATGTAG GATTCTGGA TACATATTC ATGCACTTT 1800  
CCTTCCCTCT CCTGCTTCTT CTTTCTGCG TCCCTTACT TTGCGCAAT GGGGCAACAC 1860  
CAGTCAAGTT TGTGGGGGGC CAGTCCGGCT GCGAGTTCC TGTACTACT CATTGTACTT 1920  
TTCAATTTTC CTGACCTGTC ATTTCCTAT AAGAAATTC GTGAGATGA ATTGAATTT 1980  
GTAAATTCAC CAGTGGAGC CGAGATTTTC TGGACCCCG CAGTGGAG GAGAAATTC 2040  
TCAAGCTTC CAGTGGAGCT GAGAGCGAAT GACTGTGAG CTGCGGCCCA TCACTGTGA 2100  
GACTTCCCT GACTGTGAT AAAAAATGCG GGAATGTGCG CAGAGAGCC TGAATAGGGA 2160  
TGGAAACTTA TTGTGACAA GTCTTTCAG AGGAGTTTCT TAAATGATTA TTGTATTTA 2220  
TTTCAGAGC AATTAATTC TAACTTTCA AAAAAAAAAA AAAAAAAAAA 2280  
AAAAAAAAA AAAAAAAAAA CAGGGGGGGC CCGTACCCA TTGCGCAAT ATGATCTTA 2340  
AAGATC 2346

(2) INFORMATION FOR SEQ ID NO: 194:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3054 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

TATCTGAGC ACCCTTAAT CTACATATTA TGAAGAGC TGAATATTC TAAACCCCTA 60  
AGCTGAAGT GCGCTGTGAT ACGAGCAACT GAGATTAAG AGGCTGGGC TGTGTCTTC 120  
TGTGTATGCT CTGACTCTT CCGCCAAATC TTCTCTGGA GCTTTCAGC CAGCTGTCTA 180  
AAGCAATTC GTGAGAGC TCTTCTATCT ATCTTTAA AGCAATATCT TGAATTTCA 240

425

426

TTCAACAGCT GATGCCCTAT AACCCCTGCC TCGATTCTTT CCAATTAGGC TATTAGAGCT 300  
ACGAGAGTCT TTACATAATT CAGATGGGTT TCAATTGCCCT CTAACCCCTCT CTAATGCCCC 360  
5 CTCCATTATT TTGACTTAAG CAGTCACAG TGGCACTAGC ATTATACAA GAGTATGAGA 420  
AATACAGTGC TTATATGGTC TACATTAAT GCGTTCACTA TCGAGGCTGC CTGGAGGAAG 480  
GATGGAGCC TCGGGGCTTC CTTATGTCTT CCACACAGG AGCTGCTTCA TCGAGGTCAAT 540  
10 CTTTTCCTCC TATCTCTGTC TTCCCTCCCC GCGTCTTAAT GGTATCGTGG TACCCAGGCT 600  
GGTCTCTGGG CTAGGTAGTG GGGACCAAGT TCAATTACCT CCAATCAAGT CTACGATAGT 660  
15 AAACCTAGCT ACCAGTGTAA GTGGGAGAG CTGGGTTTTC CTAGTATACC CACTGCATCC 720  
TACTCTCTAGC TGGTCAGGCC GCTGCTTCCA GGTATGGGAC CTGCTAAGTG TGGAAATTAC 780  
TGTATAGGCA GAGGGAAATA CAGGAGGGCC CTCGTGTGTT CCGTGGCTCA GCGAGCTGCC 840  
20 CACAGGCCAT AAACCAATAA AACAGAAATA CTGAGTCAAT TTTTATATCTG GGTTCCTCTC 900  
ATTCTCACTG CACTGTGTGC TCGTTTGGCT GACTGGGAAAC ACCCAATAC TACAGAGTCT 960  
25 GACAGGAGCA CTGGAGAGTG TCCACTCTTA CCTGGGAAT TACTGTGTAA ATTAACCTTTC 1020  
AGAACTGCTA CCAATGAGTG AAATGGCCAC ATTTCGCTTT ATAAATTTCTA CCAATGTGG 1080  
30 GAAAACTGCG CTTTTCCTCA GCGCTTCCA GGGCAATAAA CTCAGCCCT TCGATAGCAA 1140  
GTCCATCAGC CCAATATATT TTTTAAGAA AACTTGCACT TGTTTTCTCT TTTACAGTTA 1200  
CTTCTCTCTT GCGCCAAAT TATTAACCTT AAGTGTAAAA AAAGTCTTAA ACACAGCTT 1260  
35 CTGCTCTGTA AAATATGTA TTATACATCT GTATTTTATA ATTCCTCTCC TCGAAAATGA 1320  
CTGTCCCATT CTCACATCAC TCGATTGGG GCGTTTCCA TTGGTCTGCA TTGCTTTTAT 1380  
40 CATTGCAAGC CAGTGGACAG AGGGAGAGG GAGACAGGG GTGGGCAACA CTGTGTGTC 1440  
TTTCTGACTG ATCTGTACCA AGAAGAGTA ACAGTAGGC TCTGCTGCC ATGCACACT 1500  
CTCCAAACCA CTTATCTGCC TCGAAGAGTG GCGTTTCCAG GGTCTTTACT GGGAGCAAT 1560  
45 TAAAGCCCTT CCTACACCTT TCTTTTTC TTCTTTTACT CTTTGGCTT CAAGATTTT 1620  
TGGAAAAGAA ACAATATGCT TTACACTCAT TTTCAATTTC TAAATTTCCA GGGGATAGT 1680  
50 AAATATAGG CAGGTGGCTT AAGGCTGCTG TAAAGTTGAG GGGAGAGGAA ATCTTAAGAT 1740  
TACAGATTA AAAGAGATC CCTTAACAA AAGAGACAT AGAACTGTCT TCTCATTTG 1800  
CCACCTTTCC TGTTCATGAC AGTACTAAC CTGGAGCAG TAACTTTCA TTAAACCAAG 1860  
55 AAAGTGGTTC ACCTGAGCTC TGAAGAGCTG AGTACTCAG CCACCTCAAT CACCTTACAA 1920  
GATCCCAAG AGGTCCAGG AGTCCAGCT CTTTAACCTG AGCTTAAGCA ATTAACCTGG 1980  
GCAAGTGGG CAGAGAAAT GAGGAAGAT CCACTGTGTA GGTGACGCC AAGGATGAAA 2040  
60

GACAAAGAG GAAAGAGTA TCAAGGCGAG AAGGAGATC ATTATGTTGG GTCTGAAGG 2100  
AAAGTCTTTT GCTATCGGAC ATGTACTGCT AGTACTGTGA AGCATTTTAG GTCCAGAAAT 2160  
5 GGAAGAAAA ATCAGCTATT GGTATATATA TATCTGCTTT TCCCTGGAGT CAGTTTTTTT 2220  
AAAAAGTAA CTCCTAGTTT TTACTGTGTT AATTCATAAA CAGAGGGGAG CTGAGGCCAT 2280  
TCCCTTAGG AGTAAGATA AAGGATAGG AAAGATTCA AGCTCTTAAT AGATCAGAG 2340  
10 CTTTCCAGG TATTAACCT AAATTAAGA AGTACATTA GCGAGGTGG AATATGATCT 2400  
AGTTCTGAT AGCTACCCAC AGAGCAAGTG ATTTATTAAT TTGAAATCCA AACTACTTTC 2460  
TTATATCAC TTGTGCTCC ATTTTTCOA GGCAGGAAA TATGTCTCCC CCTAATTTTC 2520  
TTGCTTCAA AATTAATATC CAGCATCCA AGATCATTTCT AAGTAAATTT TTGCAGAGAC 2580  
20 ATCTCTCAC CCAAGTGGCT GTCTGGAGCT CACCAAGGT CACCAACAA CTGTGTGTG 2640  
AACCAACTG CTTAATCTT CTGGGGGAG GGGATTAAGT AGACTAGGAG ACCAGAGTG 2700  
AATGGGAG GTGAGGACT TCAATATGTT GCGCTGTGAG AGCTTGATTA GAGGCCAGA 2760  
25 CAGTGGCAG AAGAGAGAC TTGCCCCAG AAAAACCTGT GGGTTGTGCT AATTTCTGTC 2820  
CAGAAATAG GGTGGACAGA AGCTTGTGG GTGCATGGAG GAATTTGGAC CTGGTTATGT 2880  
TGTATATCT GCACTGTGAA TTTTGTGAT GTAAAGACA ATATCTGTGA AACTTAATGT 2940  
30 CTGTATAAT AATGAGGCTT AACACAGTA AATATTCAT AAGAAGTCAA AAAAAAATA 3000  
AAAAAACTG AGGGGGGGCC GGGTACCAA TTTTCAAAAT AAGATATGTA TTAC 3054

## (2) INFORMATION FOR SEQ ID NO: 195:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 907 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

GGCAGAGCTC GTGGCCGAAA CTTTATCTGC TCTGGCTGC CACTACTGC CTGGCCGCGG 60  
CCCTGGGCTG GGGCTGCACC AGCCTGGNG CGGGCTCCCA CAGACGCCCC CTTCCAGCA 120  
CGCTGCCAC ACCGGCAC TTCTGGCGGA AGTGTCTGCG CGTGGCGGG ACCATATGGA 180  
CGAAGGCTT TGTGCTCAC TACAGCTGG GTGAGCAGGG TGGCAGCAGC CTGTGTATCC 240  
TCTTGGCTCC TGTGTAGCA CAGACGGGCT TTCTGCTCC CAGTTGGGAC TGTGGAAATG 300  
TGTGGGTGCT GTGGTGTGCT CCAATGGCTG GTCTGCTGG GGTGGGACCT TGTGGCCAA 360  
GCAATGAAA CTGCTGCTC TGTGAGTGC GTGCTGGCT TCCGCTTCG GGGCTAGCC 420

480 TGTGAGACTG CCTGTGCTT CCACTTGA CACCTTGGG GCGAGATGG AGCTGGAC  
540 AATCTTGA GAGTCACTT TGTGACCT ATGTCTGAG GACTTTTG GAGCTCTGT  
600 GAGCAGACTG ACTTCACTG GGAATGATC GCTGAGCA GCTGGCCCC AGGACCTTTC  
660 AGGCAACA CTGACCTTT CTGACAGC TGGACTGCT GGGAAAGCTG CTGCTGGCA  
720 CTGTGGGG AGGACCTGGC TGAATGGTG GGGCAATC CTGCTCTT GCTCTGCTC  
780 ATCTCTCTG CTTTCCCGT TCTGACTG GACTGACAG CAGACCTTT TCTGAGACT  
840 GAGTGGCTGG AGTGTCAAT AAAGCAGAT GTGCTGTGG CCGAAAAA AAAAAAAA  
900 AAAAAAAA AAAAAACTG GAGGGGGGG CCGTACCA AATGCCGA TATGATGTA  
907 AACATC

## (2) INFORMATION FOR SEQ ID NO: 196:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

60 GGCAGAGGA GGCAGAGGA GTGGCAGG GAGAGAGCA GCTTATTGA CTGACACCC  
120 CCTCTGTGT CAGCAGACT CTTCGTTGG TGGAGGGGT CTGATACAG AGGAGCCAG  
180 KAGGCGAGA AGAATGGGG GAAAGCTGG CAGTGGGCG GGTTCGAGC AGCTTGGCA  
240 TGTGAGCTA GAGCTCTCC AGCCAGACT CTTGAACTT GAAATGTCA ACCGAGCCC  
300 TTGAGCAG CCGTCAAGA TGTATTAGC TTGATATAC TGTAAAGAA TATTATATC  
360 AACCTCACT ACATTGAGC TGAATCAG GACTAACCT GAAATGGGG TGTTCAGAC  
420 TTGAGAGGA TGGCAGAGC GTGCCCTGG GAGTGTGCA GCGAGCTTAT CTTCTCTGT  
480 TGGCAACTT GAGTCCAGC CTCTCGGCG CCGATGGGT GAGCCGTCG GTGTGTGT  
540 CTGTCCAGC GTGTGATCC AGCTAAAGG ACAGAGAGG AGCCCTGAG CAGACTGGA  
600 AGTGGCTGG AGAGAGCTC GAGTCTCCG AGTGTAGCC CTTGGATGG GATTCCTTG  
660 TGTGCTTAT TGTGTAGAA TGTGTAGAG GTGCCCTAT AAGAAATTA GGTCTGTAT  
720 GGTATTATA AAGACTTTT AAAAGCAAA AAAAAAAA AAAAACTGA GAGGAGGCCC  
780 GGTACCAAT TGGCCAGTA GTGAGTGTG TTGACATTC CTGGGCTGT GTTTAGACA  
840 CCGTTCAGC TGGAGAACT CTGGCTTTA CCAACTTAA TGGCTTGA GAGATCCCC

900 CTTCGCGAG CTGGCTTTA TACGAAAA ACCGCCAGC GAAATGCC TTCCAGAG  
960 TTGGCGAG CCGATAGC GAAATGCA ATGTAGAG TTATATTT TKTATAAAT  
1020 TCGATTTAA AATTGTGTT TAAATCACT CAATTGTTT AACCAATTA GSCGAATTC  
1080 CCGCAATCC CATTATTA TTCAAAAA ATAAACGAA AAGGGTTTG AATTGTTAT  
1140 TTCCCAATT TTGAAAGCA AATGCCCC TTATAAAA GTTGAGCC CAGCTCTCC  
1200 AAGGGGAAA AAAGTTTTT YTGCGGGGA AGGAGCCCC GATCACTTA AGATCCCCC  
1260 CCAACATTT TTTTGGGGG GTCCAGAG GTCCCCCTA AATCTTTT CGAAGCTTA  
1290 AGGAGATCC CCAATTAAA AATTGAGTN

## (2) INFORMATION FOR SEQ ID NO: 197:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1020 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

60 GGTGTGCTG GATGTCTGG TATGTAGTT TTACAGAG TTATGTAGC AAATGATGA  
120 GACTGTAG GAGAAATAT TGAATCAT TAAAGAGAC CTGATATG GAGAAAT  
180 SCATGAGG AATTGATAT AGCTTTAG GTATCATTT TTCACTGA GCAATGCT  
240 TTTCAGCA ACTTATCC ATAAAGGCT GTATGACAT CTATGAGTA TGACAGAC  
300 ATGTATAG TGTATTTCT TCTTGAAC ACTGTATTT CTACAGAAA TTTCAGAC  
360 ATTATAGG TACAAAAAA TCCAGCAG CTGTAAAGG TGTGATTC GATTATATG  
420 TATGTATAC TGAATCTGT AATGATCA GTTTTCTT TTATCTCT AACCAAAA  
480 TTGTATAT TTGATCCCA AATGTTTA ATCTGTAT AATTATTA AATCTTTT  
540 TGTATCAT TGCCTTTT GTGTGTGA AAGACTTAC TTGCCTTG AAGATGAT  
600 ACTCTTTC ATCTTACAA TATGTATAT GTATTTTC ATAAAGAG TGAATTTGA  
660 ACCAGATAT GTGATTTGT TTATAGAAA AAAGCTGCT TCAATCTGT GAAATCTC  
720 TTGAAATAT TCTTTTACA GTGTAGCC AGCTAGATA CCGTATGCT GTTATTTCT  
780 TTCAATAG CTGACATCT AATTAGCA CTGAGCTTT TATTATTT CTATTTTGA  
840 AGTTATTT TCTTACTA TTATATAT ATACTTTA TGTGTTTG AATGGAGCT  
900 GAACTTTG GATATGATA TTGATTTGA AATATTTTG AATTTTTCT ACTTAAAT  
960 TTGAAATAT AATGAAAT TGTATATGT ACTGAAAGT AGTTGTGA CAGTACAC

TCTCTAATAA TATGATGCTT TCCCTTAATAA GAGGAGGAC ATGTCCCACT TTCACACAG 1020

5

(2) INFORMATION FOR SEQ ID NO: 198:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 524 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

AATTCOCGAA CCTGAGGCTT GTGTGCGATC GGGCAGGCA AGTCATTTTGA CCGGACCTT 60  
CCCGGCGCAG AAGACTGAA GTTGATTTGA GAGCCTGTAT TTGGGTTTAA GCGAGGCTC 120  
TGGGGCTTTC GTCCCGGCC AGGACACAG TTACTTTGCA CCGGGCGGGC CTGTGCTTAT 180  
GATGTTCTC AACCCAGGGG CGGCTCTGTC CCTCTACTCG TCCAGAGGCC ACTTGCAGG 240  
CAGGAGCCTT CCCAAGCCT TAGAGGCTGC TCGGATCTAC CTGTTGGAT GCACTAANAG 300  
GACCTTTGTC TGGGACAGG TGTCTCAAC ACCCTGCTGC TGGCTGCCAG GCAGGCGCTC 360  
TGGAGCGAA GGGCAGGAC TCATCAGAC CTCCTCGAC CTTCCAGGCC AGCGAGTTGG 420  
CCCGAGCCA AGCATTTGGC TCTCTTCCC CCAAGGGGAC AGGAGGCTC TTGGGCTCT 480  
TCCCTTCTTG GACAAGGCC CCTGCTTTTG CCTCATATAA ACTG 524

35

(2) INFORMATION FOR SEQ ID NO: 199:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 332 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

GTGATACAG GAGGCTGAT CATCATCTGT CACCATGCAA TTCTGTGCTA CAGCCTTCT 60  
GTTGTGCCA CTTCTGGCTC TTGTGATGT CCCCATATCC CTAGGCTTCT CCCCCTCTTA 120  
GAGGGCTTC TTGATGATTT AGAATATAG AATGATGAC ATTCTCATG TCCATATAG 180  
AAGGAGCCAC AAGACATGTC TTTTAATAA AAGGACAGTG TCCATCTTT TAGCTGCCA 240  
ATAGACCTT GGTCTCATCC TCTGGAGCT AGGCTTAAA ACAGCTTCTG TTTTCTTAT 300  
TGTCTCATT GTTTTGCCAA GCTTTATTC GG 332

60

(2) INFORMATION FOR SEQ ID NO: 200:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 376 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

CCAGGGAGC CCAGGCTTG TCTGATATG ACATCATGTC TTCCCTGAC TCCTTCCCC 60  
ACCCCTGGGC ATATCCGAC GAACTTATG TTTTCTAGAA GCTAAGCAGC TCTGGGACT 120  
CAGGACTGG TCCAGGTAGG CTGATGCGCA GCTCAGTCTT AAGAGGTCTC TGAAGATCTG 180  
GACTGAGAC CTGCTACTC CCAGAGCCAG AGCCATCAG CCAGGCTTGC TTGAGGCAC 240  
CTGCTGTGG ATGCTGAGC TCACCAAG GCTGGCAGC TCTGGGCTC ATTTAAGGA 300  
TTCTGATAG CCATGGGCC CTGAGGCGAG CCATTAAG CATCTGGCTC GTTTTGGAA 360  
AAAAAAAAA AAAAA 376

(2) INFORMATION FOR SEQ ID NO: 201:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1192 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

CCCATATAT TTCTATAACA TTATTTTAG TGAATTTATA ATGTTTCTTT GTATTAAATT 60  
ATTAGATAT ATCTTTAGAT AATATGTTA CTAAATTAAT AGGTAAATTA TATTTTATTC 120  
AAAAAATAT TGTGATCTA ATGCTACCA ATTAATGTAC TTGTAGATGT ATCTATCTT 180  
AACTGAGTC TTCTGTCGC CTAAATGAGT GTGAAGGACT CTTCCTCCCT GGGGAGTTT 240  
TTCTTTTTC GGAGGGAGGA GGGCTTTCC AGGTATATGT TCTAGATGT TGGCAGAA 300  
AATCTGGAC CAGACACAC CAGTCTCTC CTAAATCCAC GTCATTTGCC TTCTATCCCA 360  
GCTATCTTC CAGTGTCTC TGGGTGTTT CAGAGCCAC AGGAATGAA TAATCTCTG 420  
ATGAGTTT TTATTTTCT TCACTTGT TTACACTGTA WTTTCTGAT TTATGGGTG 480  
CTGTGATTA AAAAAAATA GTGAATTA GTAACTCA GGTCAAGA AATATACATA 540  
AATAGATTA AGCTGACCTG TAGATTAAT CAGGTTTAA RAGCTTAGG TTCTTAAGT 600  
TGTGTGAAA KTTTCTCTG ATCTTTCTGA TCCGAGACA AAAAAAGCAG TCATGTTGT 660

WATGTCATG GAAATGAAAC GAAAGAGAG GCAATGCTG TCTTGAGGAG CAGAAAGCT 720  
TGTGTGAC AAGCTTAAAC CACCACTTTC ATGAGACATA GATTATGTC TGGAACATAT 780  
TTGACACCGG CTTGCGAGTA AACCTTGTA GTGTGTGCA GTGAAACCG TCACTTTCCG 840  
CTAAGAGAGG GCGTGTGTG CAGCGGAAAT GGTGATCTGC TGTAAACA CAGCTTCAT 900  
CGTATGTAT GCTCCCTACT CAAAGAGGT GGTCCCAAC AACCTTGGG AGTCTCTCT 960  
TGATTCATGG ATGAAACCTG GAACTTTTG AGCACTGAGT TAAACATAG TCTTAAATA 1020  
ACTCTCAGA CGTTTTCCT AGTTATGTC TAACTCAGG GTGTGACAG GCGTTTCCA 1080  
AGTCAATTT TCTGGAAT ATTTCCAGTG TTATTTGCA CTTAGGCCA CTCGTGTAG 1140  
CGTATTTCT TCTAACTCA CGATTATCT GAATAAAGT CAATTTAGG GG 1192

## (2) INFORMATION FOR SEQ ID NO: 202:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 589 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

ATCTTGAGCT ATCTTTCAG GGGATCTTT GCAAGTGAAT GCTTTCACA AGTAAATATA 60  
GTCACTCCCG AAAAGATGAG AACCTGAGTT CTCACAGAT TGAATAGAT GAGAGAAACA 120  
CAGAGTTGA GATTAAGGAA CCCATGCTTC GATTTCTCA TTCTAAATTT GTTCTTACTG 180  
AAAAATATAG TATCTGAAT ATTCACGAC AGATGTGTA AGTAAACTG AGTCAATAG 240  
ATGACAAAC AAAAGAGAGT GATTACAGCA CCGAGATGA CATTAATAT TTAGCCACTG 300  
GTTCGAAAGG CAGAGAGAA AACGTAGAG AAGATGTTTG TATTAATCTC ACTTTGATTT 360  
CCGAGAGTCA GCGAGTCCG TCAACAGCTA CTCGATCTGA GGCATTTCT AGTGTGTAG 420  
ATCAGAGGAG AGCTATGGA ATTAAAGAC AACATCAGA GAGAGGCTCT TCAAGCTCTG 480  
AGGTGAGAA ATTCCTGAG AACCTGTG AAATCTAGG AAGAGAACTC AAAAGAGAA 540  
ATATGAGAG TGTCCCTG CACTTCTC TCACTGAAC TCACTCCA 589

## (2) INFORMATION FOR SEQ ID NO: 203:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 847 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

GGAGAGAGG CAGCTGCTG GCGCCATCA AGCGTGTCC CAGGTGCGG CTGAAGACC 60  
GAACTCTGG GAAAGAGAT AACCTCAGG TGAACATCTC CAGATCTAC ATGATCTCT 120  
ATGACCTGCA GCAATCTG AGCACTCAC ACCGAGCCTT GAGAGAGAG ATTTGACAGC 180  
TGCAGGAGAA GCTGATGCC CTGACTGAGC TGTCTAGAC TGCCTTGAGG CAGAGAGCT 240  
TGCAGAACTT AGCCAGCAT CTAAATACT GAACTCAGA GAGAGAGCA GCTTACTTTC 300  
CCCAATAGTG AGTGTGAGAC ATGCTCTG CCACTCTGA CCACTCTGA CAAAGACAT 360  
CAAGTCAG AGCCAGAGG GCGTGCCTT GAATGCTTG GCTTGTGAT GCTGTCTGAA 420  
GGAGAGCTG GCTAAGTGG GAAAGCTTG GCGCACTGA GCGCCAGAT GGAACATGG 480  
TCAACCCGAC TGTCAATACC CTCAGAAA AACATCTAC TATCTGTGTA TGAAGAGCT 540  
CGAGCTTCA GTTACAGTG CAGCGAGCTG GAGCGAGAC TGTGTGCTC CTGAGAGAA 600  
GGTACTAGG GCGCCGATC CAGATCTG GAAAGCTTCA GTTACTCTG GCGAGCTGA 660  
AGAACTGGT ATGAGCTTGG GCGCGGCTG GAGGTGCGC GCGTGTGAG CAGAGAGAG 720  
AGAGAGCAT TTATCCAGAG CTCAGAGAG CAGCTGTGAG GAGAGAGAA GTTAACTCA 780  
CGAGCTCTG CTTTATCTT TGTAAATAT GTTAAAGCA GAAAAAAA 840  
AAAAAA 847

## (2) INFORMATION FOR SEQ ID NO: 204:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 852 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

AGAAAGTAC TGCAGAGAG GATGTCAAG CTGCCCCAG CATGACGCA AGCTGTGCC 60  
GCCATAGG CGTCCCGCA GATGCGCTG AAACACGGA AGCTCCGGA ACAGATGAC 120  
TCAATGTGG AACATCTGA GATGCAATG ATCTGTATG ACTTCAGCA GAATCTGAC 180  
AGCTCAGCC GCGCCTTGA GAACAGATT GACAGCTGG GCGGAAACT GAACTCCCTG 240  
ACTGAGCTG TTAGAGCTG CTTGGGCTGG AGCGAGCTTC CAGAACCGG CCAAGAGTCC 300  
AAATAGCTG ACCGAGAGG GAGAGAGAG GATATTTTC CCAATCTGA GGTGTGAGC 360

ATNGCTCTCT TGGCACTCCN TGNACCCAGC CTTGACGAAA GCACCTCAGG TGCAGAGACC 420  
AAGGGGGGCC CTGGCTTGGG GTGGGTGGC TTGCTGATGG CTGCTGGAGG GAGAGCTGGC 480  
5 TAAAGTGGGK AGGGCTTGGC CCACCTGAGG CCCACAGTGG GACACATGTC ACCCCACATC 540  
TGCATACCTT CATCAAAAAC ACTCTCACTA TCTCTGCTATG GACGACCTCC AGCTCTCATG 600  
TACAAATGCA GGCACCTGGA GGCAGACTC CTGGGTCCCT GGGAAAGAGG GTACTAGGGG 660  
10 CCCGATTCOA GGAATCTGGG AGGCTTCAAT TACCGCTGGC CGAGCTGAGG AACTGGGTAT 720  
GAGCTGGGG CGGGCTGGA GTTGGGGCCC CTTGGTGGGA CACCAAGAGG GACACATTTT 780  
15 TTCCAGAGCT CGAGAGGCA CTTGGTGGGG AGGAAGAGT GTAACTCAAC AGCCTCTGCT 840  
CTTATCTTTG TA 852

20

(2) INFORMATION FOR SEQ ID NO: 205:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1354 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

(2) INFORMATION FOR SEQ ID NO: 206:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1378 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30

GATTGGGCAC GAGGCTTGGT GGAGCAGAG AAGTCTCTTG CGGCTGGGC ACTGTGCTG 60  
35 GCAAGACTG GCATGAGACT CATGTGCTG CATGCAAGAG TCTGTGTGTT CGGGGGGTGC 120  
TGGCTGTGTA ATGCCACTGG GCACCTTTCA GACACACTTT GGTGTGATCC CATCACTTC 180  
CTGACCATCG GCTATGTGTA GTGTGTGGG GGCACCATGT GGGGCAGAT GTTGTGCTG 240  
40 TGCACCTGAG TCATGGGTGT CTGTGCAACA GCGCTGCTGG TGCCCTGTGT GGCCTGGAG 300  
CTGGAGTTTA ACNAGGCA GAAGCAGTG CACACTTCA TGAATGATAT CCAATATACC 360  
45 AAGAGATGA AGAGTGGCC TGCCGAGTG CTACAGAGG CTTGTGATTT CTACAGAT 420  
ACTGCAAGGA AGGAGTCTCA TGTCTGGGG AGGCATCAGC GCAACTGCT GCGCGCATC 480  
AAGCGTTCC GCGAGTGGG GTTGAACAC GGGAGGTCC GGGACAGT GACTCCATG 540  
50 GTGGACATCT CCAGATGCA CATGATCTG TATGACCTG TATGACCTG GAGCAGCTCA 600  
CACCGGGCCC TGGAGAAACA GATTGACAG CTGGGCGGGA AGCTGATGC CTTGACTGAG 660  
CTGCTTAGCA CTGGCTTGG GCGAGGCGAG CTTGCAGAC CGACGAGCA GTCCAGTAG 720  
55 CTGGACCTAC GAGGAGGAC CAGGCTACTT TCCCACTGAC TGAAGTGGTG GACATGCTCT 780  
CTGCCACTCG TGACCCAGC CTTGAACAA GCACTCTCAG TGCAGGACC AAGGGGGCC 840  
60 CTGGCTTGGG GTGGGTGGG TTGCTGATGG CTGCTGAGG GAGGCTGGC TAAAGTGGGK 900

AGGCTTGGC CCACCTGAGG CCCAGGTGG GACATGCTC ACCCCACTC TGCATACCTT 960  
5 CATTCAAAAAC ACTCTCACTA TGTCTGCTATG GACGACCTCC AGCTCTCAAT TACAAGTCCA 1020  
GGGCACTGGA GGCAGGACTC TTGGGTCCCT GGGAAAGAGG GTACTAGGGG CCCGATTCOA 1080  
GGATTCGGG AGGCTTCAAT TACCGCTGGC CGAGCTGAGG AACTGGGTAT GAGGCTTGGG 1140  
10 CGGGCTGGA GTTGGGGCCC CTTGGTGGGA CACCAAGAG GACACATTTT TTCCAGAGCT 1200  
CGAGAGGCA CTTGGTGGGG AGGAAGAGT GTAACTCAAC AGCCTCTGCT CTTATCTTTG 1260  
15 TAAATATGTT TAAAGCCAGA AAAAAATAAA AAAAAAACTCG AGGGGGGGCC 1320  
AGACCAATC TGCATATAGT AAGGCGCCNN AUAU 1354

20

(2) INFORMATION FOR SEQ ID NO: 206:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1378 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

TCCCAAGTG CACAGCCAGG GCGCTCTGTT CTGCAAGAGA ATTCACAGCT GGTGTGGAC 60  
TCAGCCCTTA GNCATTCAA AGCCTTAATG TTGTATCAT ATCTTACGTG TTGAGAGCTT 120  
35 GACTGAGAA ACAAAATGTG CATTAACGTG AATTTATCT TAGAGATCTG TGCAGCTTAT 180  
TTCTGTACA AAGTTATAT TGTCTAATA GAGAATCTT AATGGCTCTT GTGATTAATG 240  
TAACTCCAGT TACAGGTGA CTTTAAATAG CATACAGTGA TTGTATGAA GGAAGTCAAA 300  
CAATGTGGG ATGTGTGGA AAGTTATCTT TCCGCTCTT TGCCTGTGTC ATTGTGCTT 360  
GCGAAGGA TGGCCTGAT GCGAGGAG GCGAGCTGT AATTAATAA TAATTCACAC 420  
45 TATGAGACTA CGAGGCACT AGACTGGA AAGACCAAG AACAACAG ATCCAAACCT 480  
TTCATCTTAC AGGTGACAA ACTGTGATGA TGCACATGTA TGTGTTTTT AAGCTGTAG 540  
CAGCTTAACA AATGTAAAT TTGCAATAT TAGGAAGTGC TGGTGGCAT GAAGAAGCAC 600  
CGAGGCACT TGAATCCAG TGTGTGCCC TGTCTAACC AGACACACA GAGGCTGGGT 660  
720 CAGTTTCCC TCAGCTGCTT AACAAAGTTC CTGCAACAGA AAGTCTTTAC AAGCTGCTCT 780  
TCTCGATAC TGAAGGTGC AGTTTTCTGA ACTGACTGA TTTTATGCA GTTGAATAA 840  
AAAAAGCT ATTCACAGA TTTCAGAGG TTCTGAGACA TCTTCTGATG GCTTTACTTC 840  
CTGAGGCA ATGTTTTAC TTTATGATA ATTCTTGT GCCAGGAT AAGGTGAGA 900

AAAGACACT TTTAATATAT AGGTCTCTCT GAGAGAGAC TAAATTTAGA AAGAAATCT 960  
GTGACAAATTT TCAATATCTC ATTCTTAAA AACCTATATC TTAACTTACA AAAGTCTTTT 1020  
TGAAGATAGG TTAGACACAA TGGACAGAC AGTTGTCTCT TAAATGATTA GTGCTATATC 1080  
TCAATATATC GATTACTATAC TACTGCTCTT AAAAAAAAA ACCAGCAATAT TTAATGAAA 1140  
CATGAGACAG GATTATAGTGT CCTTAAACCA TATATTTTGT GACTTAAAA ATATCTTTAA 1200  
AACTGCTCTT CTGCTCTTGT ACCATCTTAA GTGCATATGA TTAATTTCTAT GTACAACTGA 1260  
TCTGTCTCT TATTTTAAAT AATTTATCAG ACTGAAAAA AAAAAAAAAA AAAAAAAAA 1320  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAA 1378

20 (2) INFORMATION FOR SEQ ID NO: 207:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1166 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(\*) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

AAATCACTGC AATTAAACC CCTTCCCTCT CAGAAATGTT CAGAACCCGC CATGATATAC 60  
CCTCAATTTA GATGACCCTC AATATTTAG ATGACTGTGT GATCCCAAG AGCTAACCTTT 120  
GAAAGGCGCA CTCAGAAATA AAGATCTTGA CATTCGAAA CATCTGAGC TTGTGACAA 180  
CATGATATGC GCTCATAGAG AATGCGGCTT TCGAGAGACA AATTAATTTAT GATATGTGAG 240  
ATGCTATGTA CCTGTGTGT CATGTCTCTC GTGTGTGAGT ACTTGACATG AACCAAGCAC 300  
GCTCAATGCG TGAACAGCAT TCCACAGCC TGCAGATGTG TGTGTGTGT AATGAGAGAG 360  
GGGAGCCGAG GCGCGGCTTT TGAATGTGTT GCGTGTCTGA ACTGTAGAGA CACTTGTGAG 420  
TGAATGTGT CTAATTTTCA AACTGTCTCT TTGTCTGTGA CATCTGAG AGGAGACTTAC 480  
TGCACAGACC ATGCGCGGTG CTTACAGAAAT GAAAGAAATC CCGGTGTGT CTGTCTGACT 540  
CTGTCTGTGA ATGCGGAGAG GAAAGAAATG CTTTGTGTGC TTGTGTGACA CAGCTGATGC 600  
GTGCGCTGAG AAGGTGTCCA CAGTGAAGCC TGTGTGAG AGCTGTCAAT ACCGTGTACA 660  
CCTGTGACC ATGCGGCTTT TGTGTCTCTT GATATGATGT AGCAAAATGT GAAAGGAAAAG 720  
GAAAGAGGCT TTGTGTGACA GCGAATATAT TAAATATAGT GTGCATATAC ATCTGTGTGC 780  
TCTGTGTGAG CCTTCCCTGT GCTTCAATGA CATATGATG ACTGACTGCC AATTAATGTC 840  
AACATTTCTT GAGAGAGCT ACTTATGAGG AAGCAAGTGT AGTGTATGT CCGATATACA 900  
GTAAATTTT CAGATATACA GCTGTGTGTT TGTCTTTTCT AATGTAGAGC CATGTATATT 960

TCCCGAGAGT GTTCAATCTT GACCTTATGC ACTGATTTTT TGTATGTGTT AATATAGTGT 1020  
TTGGGCTTTT AAGGTTCAGA GACTGTGAGC TTGGGCACTT GCGCCGAGAG GTTGTGTGAG 1080  
GCGCTTGGC CCTTATGAAA GTTACTTTTA GGGGCAAGAA TTGTGTGATT TTGCCCATTA 1140  
CATCTTTCAG CTGAGAGCTT TTAATA 1166

10 (2) INFORMATION FOR SEQ ID NO: 208:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 697 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

20 (\*) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

TACTTTCAG AATTATAGGA ATTAACATTT AGATGACATT TCAATTTAG AAGAAATTA 60  
GTGCTTTCA GTGCTTTTA TTTGATTTCT GAGAGAGACA GACTGTGAGS AACATTTATC 120  
CCAGGCTGT ATATGAGAGT AATCTTCAGA GCGAGAGCC AGCAAAATC AGCAATCTTA 180  
GAAAGTTACA ATTGAAAGT TTCTGTCCAG CTGTGGAGAT GACATGTCAA AGCTGATGCC 240  
AAGAACTGCC AAGATATTC TCTCATATAC TCTCTTACCC AACCAATTC AGCTTCCGCA 300  
ATTATGAGT GAGATGACT AATCTCTTT ACTTTATCA TTAATGTGA TGCATTTGAC 360  
AAAACTTTC GACTTTCGA TATTAAGGCT GTGTGTCTCT GTGTGCTCTT GAAATAGAG 420  
CATCAATAT ATGTGAGAC ATGCATATTA TCGAATTTT TCAATCTCT CCGAGAGCTC 480  
CTGAGTTAGA AATTCAGAG TTCTGAGAGT GATTCATATC ATGCTAAAT TTGAGAACCA 540  
TTGATTAAG TTAATGCAAT AAGAGAGAT TGAATAGAGA TGTGTGCTTA TCTTCTTACA 600  
GTGTGCTCT TACAGAAAA GTGCAGATTC AGTTGATGAG AATTAAGATT AATTAATGCT 660  
AAATCTTAAA AATATTTAAA AAATCTGAG GGGGGCC 697

45 (2) INFORMATION FOR SEQ ID NO: 209:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 912 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

50 (\*) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

CTGAGTACAC CTCCTATAG TGGCGGTGCG CAGAGCGCGG GTGAGCTTGC GAAAGCTTCC 60

5 GCTGCCAGAA ACCATGTCTA AGGTAAATTAA AAGGTGCGTG GGGCGAGCCA GCGTCGACCTT 120  
GCTCAGCTTC AAGGTCTATG CAGCAGCAAA AAGGAGCTCA CCGTCCAAA ATTGCGTBA 180  
GCTTGATGAG CTCTTCACTGT ACTGATGTC TGAGGCTCAA TCGAAGTATG TCGAGGAGCC 240  
AAGGAGCCAG CTTGAGAAA GCATCTACA GCTCCGACAC TATTGCGAGC CATACAGCAC 300  
CTGCTGTGAG GAAGCTACTT CCCAACTTAA CCCGAGATG CAAGTGTTCG TTCAATGCGG 360  
GTTAGACAGC TATGACTATC TCCAAATGCG ACCTGCTGGA TTTTTCGGA GACTGTGTGT 420  
TATTGCTTTT GCTGCGCTTA TTGAGTCTCT TTGCGCTAGA GGTTCAAAA TAAAGANGCT 480  
AGCTATGCG CCGTGTTTCA TGGATATGAG TCGCTGCTTC TATTATCCAC AACAAAGCCT 540  
CGCTTTTCC CAGGTGAGTG GGGAGGANTT ATATGACTCG GGTTHACGAG GATATATGCT 600  
CATAGAAGAT TTGTGAGAG AGAAGTTTCA AAGCCAGGA AATGTGAGA ATTCACTGCG 660  
AACTAAGTAG AAATCTCAT GTTCTGCCAT CTTAATCAGT TATGTGTAAA CATTGGAAC 720  
TCCATAGAT AAATCAGTAT TTCTACAGAA AATGCGCATA GAGTCACTA TTGATGTAT 780  
TAAATTTGCT TTCTCTTCTA GGAATACTA GACCAGACT CTGTATATCT CTGTGAATC 840  
ATCTACAG CAAGTAACC TGGAAATGCT TCACTAGAG ATAAATGACA AGCCTTAGAA 900  
CTCTCATTC TCAATGTGCT ATTATGTAC CT 932

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 661 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 210:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

5 GTCATCTTTT AATAAAGC TTCTGTGTTT AAGCTTTTC AAGGAGAGC ACCACCTTCA 60  
AGATTCGCC TAGGTTGAT ATGTGCTTAA TTCAATTTAT AAAATTTAT CTGTCTTCA 120  
TTTAAAGCT TTGCGTATAT AGTCAGAAAT GTCCATAATA ACAATATTT TTGTATTTAA 180  
TTTAGGAG ACTAAGGGA AGAANAATCA AAATCTAGTC TTTATGTAG CTCGAGCAT 240  
ATTAGGCGTT AAGGCGCTTT TCTAGTTTAA TGAAGATTTG TACTACTGAT TTTTATAT 300  
TCTGTGTTTT GAGATGACA GATCTCTGG GAATTCGTTG AGTTACAGT GCATTCAC 360  
GTGATCCCTC TCAAGCTCAG ATCAGTTCTA TAACCCAATG ACAACCTGTC TCTTTGTTT 420  
ACTGTCTGCT GAATGTGAG CTCAGTTTC CCAGAGTCG TGTGTTTAT ATGATGACA 480

5 GTCTTTTCC TCGTGGGAC AGTTGCTGC CCGTTAAAT TTGGGTATG TCGTTCGAG 540  
TATCTAAAC TCAAGTCTGA TGTGTATATG CTATCTCTAC TTTTAATTT ATTAATGAT 600  
5 ATGTGATTA TCTTCTTGA AGGTCADAC TTTTCAATTT GATAGAAATA AGTTTITTT 660  
C

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 592 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 211:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

5 GAACTGACA TTGTTAACA CACTAACA GAACTACTTA CCGTTGAG ATTAAATATA 60  
TAACTGTGA CATGATCAT GTACATGAT GGAATGACCA GATGCTTATG GTCTACATTT 120  
TCTTTATCC TGTATGATTT ACCTTGCTTA ATCTTGTTC CTTAATGCG TAAATTCCTC 180  
TTCACTGTTT ATTTCTAGT GACAGATCG TACATTTCT TACACCTCG CAGAGGGAG 240  
AGAAATGCT TTGCGGTG GTAACTAAT TTTTCACTGA AATATCATTA GATCAGATG 300  
GAAGAGGGA GACAAAGA GTTATACAA AAAAATGAT GTTTTITTAG CCAATTTGACT 360  
GCGCTTTTAA ATACTCTACA AGCAATTCAC GTTAAATATC ACTTTAGTG AAATAAATG 420  
TCCCATCTA GTATGTGCTT CAAGAGCCA AATGTGCTTT AGTGCCCTTA GCGTAATTT 480  
TGGTCATTTG ACATCAGAGA TGTGTAACT ATTCACCTTA ATAGCACTT ATTTCCTAAT 540  
40 AGTGTATTT TTTTGGCTAG CATTTNCTTT ACGACTAAC TTGTTGGATA GC 592

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 938 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 212:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

55 TCGAGTGCT TTCCAGCTGA ATGAATCTTA TGTCTGCGCT GCAGGTGCTT GGTITTTCAAT 60  
GTCTTCTCA ATTTTITTC TATTGGCTCT TGGAGATTN CTTTGTITTC TCGTGTGTTT 120  
GCCAGCTTT AATAAACA GCGCGAACA AAACCATAG CATTCGAA CAATAGGGGG 180



CCGACCTTGG ACCGAGTATG TCACTTTAAT GACATCGAG AAAAATCTG AATGGAAA 240  
TGAACATGAG AATTATTAAT CCGACATTT TATGCCATAT AATGGTGTG TTTCTTAATT 300  
TTGTCTCTTG TGGGAAATG TGGCTTTGCA ATTAAATGCA CCTTTCTTC TTGGAACATT 360  
TTTGTTTTGA CTGTATTAAT TAAAGTTTGG GAAAGATTCA TAAATTGAG AGAGTTTGC 420  
AACGAGGAGA TGAAGAGAG TGTGAGTACT AATCTGTTC ATGTGCTTT AGAGCAGCT 480  
ACATTTAAGT AATATTAAT TACGAAATT AATATGTGT GTATGTGCT CTACATATA 540  
AAGTACATGC CTCGACATTA TGGGTTGCTG TGCATCTGG CAATTACTGG CGAGTCTCT 600  
TTATGACAGG CAGCAGGAA CGATAGAGG GTCTGCTTT CAGAAATGC CTTCATCTT 660  
TCCGTGAAC TTATTTTCTT AATATGTGT TTTCTTGCA TTTGTGTAG CTCAGAGAC 720  
CATTTGAGC AATGTGATGC CTCATTTGCA TGAATGTAG CTGTGTGTTA AATGAATA 780  
CATTTTCAAT GAGAGCTTG ATGACTTCTT ACGTTTAT TTTTCTGTGA AGCTCAATG 840  
GCTGAAACCA AACAGGCTT TTAAAAACT GTGTAGAGA AAACGAAA AATCTGTGTG 900  
GATGTCTTT CCTGTGAA CTCATTAAA ATTCTTT 938

30 (2) INFORMATION FOR SEQ ID NO: 213:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1079 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

AGCTTACCGG GAGATGTGTG GCATCTTAAA GGTGTGTGT GACTGTGTG TGGGGAGAT 60  
GGGAGCTGTT TTAACTGTG TCCCCCTTC CTGTGCTGAC GTGGGCAATC CCGGGGAGAG 120  
TGAAGCGCG GGCCTCTTC ACCTTCCGAG TCGAGCGAGC CTGGGCGCGG GCGCGCCCCC 180  
CGAGCAGCC GAGGATCTCG TTCTCTCTTG GTTACGTGGA CTGTGAGACT GATCTCTTGT 240  
GACTGAGGC CTGTGAGAGG TTGAAGGTGA CCTGCGAGG TCGACAGAG GCGGTGAGGA 300  
GAGAGAGAA GGCATGAGC CGAGTTGAG GAAATCGTGT TCGAACTCT ACATCTGAGG 360  
ATTCATGTGG GAACTGTGT GGCATGAGG TGGGGCAGAT GTCTTGAGG TTCTACAGA 420  
AGAGAGATC TGCCTGACCA TTCTGAGAG AGTGCATCC ATGGGAGTGG TGGAGGTGA 480  
AGGTGCATGT GTTACCTCTG GCGACGAGAC AGAGCGGCA GATCTGCGCG GAGAGGTGG 540  
GTGAGAACT CTGGGAGAG ATCATCTACA TGTGTGAGGT GATGAAATGG CATGTGAACT 600  
TCCCGAATAT GCGGACAGAG TCGAGGTGG ATTAAGTGT TTGACAGAGC TTGCGGAGAG 660

TGCAAGCTTA CATTGACAG ATCTCTTTC AGATCACTGA TGGCTGGGC ACCTGATGA 720  
CGACGACAT GCGGAGGCTC ATCAAGAGCA CCGTGGCTT CTGAGCTGG CTGAATCTT 780  
GGGAGCTCT TGAATGCTTC CAGAGCTTGG CTTTTGGGAA TTGCATTTT GGCCTTTGG 840  
GCTTGAGAC GTGCTGTGGG TCAATGTGGA GACTTGAGG GGGCAGCCCC CGCTGCTTC 900  
TTGGTTTGT GATTGCGAGC CTCAGTCAAT CCTTTAATC TTGTGTAGG GTTCAGTCTT 960  
GCTGTATG TGTCTGATTA GCGCTGTGG GATTCGCTCT CTTTCTGAC TGTAGAGAG 1020  
AGCGACACT GGAATCGGGA ATTAAATTGA GAAATGAGT TTGGCTGAA AAAAAAAAA 1079

15 (2) INFORMATION FOR SEQ ID NO: 214:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3791 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

TGAAGAGGC GATCTTGTGT CCGCGCGGCG CCGTGCATTC CTGTGAGAA GCGCGCGCG 60  
AGCGACATC ATGCTGTGGC ACTTACAGGA AGGCTTGGGC TGGTGTGTA CGAACCAAT 120  
CGACGATTA TTGACAGAG AATGACCC CTTCGAGTGG CTGAGAGAG CAGAGAGCA 180  
GAAAAAGAA GCGCGCGGCG GCGCGCTTGG GCGCGCTTGG GCGAGAGCG CATGAGGCGC 240  
GCGCGCGAG CGACTGCA GCGCGAGGC AAACATCTG CGAGAGATC CAGAAAGAC 300  
CGAGAGATC GCTGCGCCC CAGCTTGGC GTGTGTGCA AGAAAGAGCA GACGAGCGC 360  
CCGTGCGGC TTTAAGAGAG AAGGATTAAG AGGATTTGGA AGAGAGCTG ATCGACACT 420  
TGAAGGTGA GGGAAATTA TTGATGAGG AGCAGAGAG CGACAGCTC GTGAGAGAG 480  
ATTGAGAGG CGACTTGAAG AAAAGGTGGA AGAGGCGGA TTTTCAGTGG ATTAGCAAT 540  
TATTGACGA CCTATTGAG GTCTGTGGG TCTTGAAGA GTTCAGAGG GCGGTGAGC 600  
TGAATGGGC CAGAGAGATG GATTGATTC TGTGTGCAA CTGAATTTG ATTAGCATAG 660  
TGAAGTGAAT AATCTTCTT TTTCATTA CAGTGGCTTG AACAGAGAG ACAATCTGG 720  
AGTTAGGGA TGTGACACT GGGAGACTT CAAAGAGAA TTAACTGACT TGAATCATC 780  
AATGTGACT GAGGAAAGC CTGAAGTGA AGAGATCAT CAGTGGGAG ACATGAGAA 840  
TGAAGAAAT GAGTGTGAG AGTTAAAGG GAGGCTTCA AAAGAGTGA CTTTGATGA 900  
GTGAGAGCT ATTCAAAAT AGGACCGGC AAAAGTGAAT TTAAATATTC GAAACGAA 960

1020 TGAAGTGGCT GATGGGAGT GGAGAGGGG ATTGTGCTTT CATTAATCAA AGAGTCAGAA  
1080 GGCTCATGCT GAAGNTTGG TTATGGACCA TCATTTTGGG AAGCCAGCAA ATGATATAC  
1140 GTCTCAGCTG GAGATCAATT TTGAGAGCTT TGGCGGCCCA GGAGCTGGCG GGAGGGGAGG  
1200 ACGAGTGGGA GGTGGGGTGG GTGGGGGCCC AAGCGTGGC AGCAGGAGCG AAGAGTCAGG  
1260 TGGTCTGCT CCTGATGGG ATGACCCAGA GGCAATTCOCA GCTCTGGCTT AAGTCGATGC  
1320 CATAGACAA CCGTGGTCC TTGTGAGCC GTTCTGTCCA AAGCTTTTGG ATGCTTAGG  
1380 ATTCCAAAG ACTAGAAAT TAAAAAATA AAGACTGTCA TTCTATACAT TCACACCTAA  
1440 AGACTGAAT TTATCTGTTT TAAAAATGA CTCTCCCGC TACACAGAG TAACAATAT  
1500 GGTAGTCAAT TTGTATTTA GAATGTATTT GGTAGCAGGG ATGTTTTCAT ATTTTTCAGA  
1560 GATTATGAT TCTTCATCAA TACTTTTGTG TTGCTGCTTG CAATATGCA TTTCACAACT  
1620 TGAATATAG GTGTGACAG TGTGTACAG TTTAAGCTT TCACCTCAT TGTGTTTTT  
1680 AATTAAAGT TTAGAGTTC CCCCAATTAC AACTGTGTTT TAATATTTGG ACATCTGCT  
1740 TTTAATACCT GCTTTCATA TTACACATG GTCAACTGGG ACATGTAAA CTTTCATTTG  
1800 TCAATTTTA TGTGTGTGG AATCTACTT ATATGTATTT TAACTTAGTT TTAATATTTT  
1860 CATTTTGGG GAAATATCTT TTTTCACTC TCTGTATGCG TGTATATAT ATATCTTAA  
1920 TCTTTATATA CAGAAATATC AGTACTTCAA CAATTTCAA GCACATTTGG TTATTAAGC  
1980 CTTCCTCCTT GCATGGCTCA TTAGTTTCAA ATTAATACAG ATTACATTT TCAGCTATAT  
2040 TTACTTTTTA AATGCTTGAG TTTCCTATTT TAAATCTTAA ACTAGATC TTATTTGGTG  
2100 AAGTGTGTT AACTACTTAA TTGTGTGGG GCACATGCG TCAAGTGAG TAGTTTTATA  
2160 GGTATGGGT TTTTCTCCC CTTCACCAG GTGGGTGGA TAAGTTCAT TTGGCAATGT  
2220 GTATATTTA AACTGTCTG TAAATTAAGT GTCTGGCCAT TTGGTATGAT TTCTGTGTGT  
2280 GAAAGTCCC AAATCAAAA TGTATATCC ATATATCAGC ACATTTTAC CTTCTCTGT  
2340 TCTAAAAAATA AATCCAAAG GCGCTGGTG GTAGGGTGA GTGGGGAGT ATTTATTTT  
2400 TTGGATTTG GGAACAGAC AGCTTACTT TGTAAAGTTG GAACAGCAG ACTATACATG  
2460 AATATTAAC CAATAACCTT TACTGTCTCT AATTTTCTTA GATGCTATAT ATTGTGTGT  
2520 AAGTTGAGTA TTCCACAGA AGTGTATTT ATCTCTTCTC TCTTCTCCA TTAGAAATTT  
2580 AGGTAAATTA TGGATTTCTA TAAATGGAG ATCACCACTT ATTAACACAC ACATAGATG  
2640 ATCAATTTAA AAGTTTCTCT AGATGTCTT TTTATTTCTG CACATTTAT GATTAAGAT  
2700 GAAGGAATTT TTAATAATTT TTAAGATTT GTTTGTACAG TCAATTTTAG AATGTCTTA  
2760 CCTGTATATG GTATATGCA GTTTTAAAA TATTTGACAT CTTCAATCTT AATCATTTCT

2820 ATTATGCTGA TTGTTCTCA CATATACTTC TAAAGAAAC TTATATGTTA TAAGAGTTAC  
2880 TTTTTCGATA AGATTATTA ATCTCAGTTA CCTACTATTC TCACATTTTA GGAGAGAGGT  
2940 AATGTTTTT AATGATGAT AACTGTGTG TGTGTTTTTG GATCTTATGA TGTCTGACAT  
3000 GTTCTGCACT GGTGCTAATG TCTATATATA TTATATATTT ACACACATAC GTCTTACCCA  
3060 GAGATTAAAT TAGTCCATAT GAATATATGA CCAATGTGTC ATTGAGACAG CAACATAGCC  
3120 ACTCTTAAT CAGTGTGTT AGACTTTTCA AGATATCTAAC TCAATTTCCA ACATGTACCA  
3180 TGTTTTATTA ACCTCTTGAT TTCCAGCAAC ATACTATAGA AAGCACTGCG TACTCCAAAC  
3240 ACACTTCTC AGTGTCAATC ATTGCTGTG TGAAGACAA CATAGCAATA TCTGTATGAT  
3300 TCAAGCTTT CAAGATAGCC TBAACTTAAA AAGTGTGTGC ATTATTTGTA TCTGATGAT  
3360 ATAAATTTG CTTCTAGTTC ACTTGTGTG AAGAGCTTAA ACTGTGAGC TAACTTTCTC  
3420 TTATTTGGG GTATTAAGT AATAAAGA TTATTTTCCA TGTCTACTTC TTAATAAGTCA  
3480 TAAAAAAT CAATAGGAT CATGTTTAT GTCAATGTGT TCTGTGKTC TGAATGTGT  
3540 GCACACCTT GTGTGTTTAT ATTTTAAA TTGAATTTTA TATGGGTTT TTATTTCTTA  
3600 AAGACAGCG TGTGATACA CATTTGGGA GGTACTTAT CTTAATGACT AATGACTTAA  
3660 TTGGAAAGT TGAATTTCTG TAAATATCAA AATCCAGGA CTTCTTTGGA TTAATCTAA  
3720 TTGTACTTC NTAGGCGA TNCATTTT TGTATATGG AAGTTTAGC ATTCGAGAT  
3780 CTACTTTTG TTGACAAAG GGCCTATAG TCCGGGGGA AATCCCTAAC NGTAAAGAT  
3791 CCAAGTATG G

(2) INFORMATION FOR SEQ ID NO: 215:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1334 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

	GGTGTCTGGGCG AAGAGACTGT CCAAGAGAC CAGAGAGAC TCGAGACAA	420
	AACAGAGATC ACCCTGAGAG GGTGCGGAG AACAGTTGAC AACCTTAAC GGGTCTTCAA	480
5	GGTGTAGAG GAAATGCGGG GGTGCTGTGT GAGCAATATT GAGCAAGCT TGTCTCTGTC	540
	TGAGCCGTGT GCGAGGGACT ATGAGAGCAT GGTCTCTGTT GCGAACAGC GCTTTGAGAC	600
10	AGGAGAGAAA AATCTGCACT ATCTGAGCTT CGGTGACTTT GGGTCTGCG CTGAGCAAT	660
	GATTCAGAAC TGAACCTTGT GAGCCCTGCA CTGACAGATG GATGACATGG AACATGACATT	720
15	AGAGAGAGAA TTTCTTCAGAG ACTTGAGAGG GCTGAGAGTG CTAGTGGCTTG AACAGAGACT	780
	TCTGAGCTTG CAGAGAGAGC TGTGTGAC TCCTGTGCGG GAGAGAGTGG GGGTCTCTCTC	840
20	TGAGATGAAA GCGCACTTCA AAGAACTGTTC CGGAGGAGCTG GTGAACTGCG CGGCAAGCT	900
	GACCCAGAAAT AAGAGATGCA GAGAGCTGTT TGTGAGATC TGAGAGAGAT TTTGTGAGAAC	960
	CTGCGCGCTCC GAGCACTGGC GACTGAGCGA CGTGCAGGTTC TTCTTGAAATC AGTATTCAGC	1020
25	GTCTGTCAAC TCCCTGAGTG GGTTCGAGCA CGAGAGCTCT GGAACAGTAA GATGAGAGCC	1080
	CTGCCCGGCT GCGTCTGCGG CGTGTATCAT GACTGAGATG CGTCCAGAGG CTGCGCGGAC	1140
30	GCTGAGAAAT AAGTTGACTT GAGTTTGAAG ACTGTATCTC GCTGCCGAGGA GGAATGAGGG	1200
	GGCGTGCAGG TGTGCTGTG TGCTCTCTTG AACAGCTGAT GTGCGGTGAC AAGATGAGAT	1260
	GTGTGCAATG GCTGCTGTGG AACTGAGACA TATCTGAGGG AATGTGTCTT GTCTGAGGCC	1320
35	CATTCAGCAG AAGA	1334
40	(2) INFORMATION FOR SEQ ID NO: 216:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1511 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
45	(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 216:	
	GTGCGCGAGG TGTGCGAGGG GGGTGTCTTG CCGCAGGGGG GCGCGCTGCC TACCTTCAGG	60
50	ACTGTGCGCT AATGCTGCAG GGGTGTTAAC CGCAGAGCTG GTGTGATGCA CTTTGAGCGG	120
	CAGAGAGTGA TGGCTGTGAC TGAATATATC CCGCGAGAAC GAGCGATCCA CCGATCATGC	180
55	CTGCCATGTC CTGCCAGGCG CGCAGAGAGG GAGATAGGCC TCAATCAGGT TTTGCGCGGG	240
	GAGATGAGAG CAGTTTTCAG GAGAGAGGCA ATGATAGCGG TTGTCCAGAA TGTGAGCTTG	300
60	AGTGCAGAGG AACAGCTTCT TATGTGACAC GAGCTGCAGG AACAGAGAT CTTATGAGAG	360

	ATGCTGCCCA AACAGTCTCT GAGGCGCTTC CTGAGAGGATT CGAATGACAA AATGTGCTG	420
	CGCGTTTTC TGGGCGAGAA CATCTGCTG TGCAATGAGG AACCGAGAT CAGAGAGAGC	480
5	GTACAGATCT TAAAGACTGT GCGATTCCTG CGGCTGCTAG GTGGCTGAT TGAATGACAC	540
	ATTCCTAGCA GCGAGGGCTT TATCAGATAC TGCAGAGCTCC CGAGCTGCC CGTGTGTGAG	600
10	GCGAGAGCTG TAAAGAGCTT GAGTGCCTTC AACAGCGAGG CCGAGCTGCT GCTGTGACAC	660
	GAGCGCTGCC AACTGACAC CGTGTGTGAC GAGTACATCA GAGAGCAGG CAGAGAGGAT	720
	TGTGTATGT GGGCGAATGG GAGCCAGAT CCGTACAGTG TTGCGAGATC GTAGCGAGGCC	780
15	TGTTTACCA GCGCTGCCCA TAAATGACAT CTGCGTTATT GGTGTGCTAT TCCCTCATGG	840
	GACATGCGAA AAGACTTGGG GTGCGGAGAT GTGTTTGCA CTGCGTTTTC ACTGATATGG	900
20	AATATGTGAG GTATAGGGCC ACTTGAGAT GAGAGAGATT CGATTGAGG TGTGATGACAC	960
	GCGCTGCTCT CTTAGTTTTC CGAAGTGGG AGGTGATAG AACAAAGTCT CTCCATTTCTC	1020
	GAGTGCAGG GAGAGATTC TGAATGATTA GGGCTATGAT CGGCTGCTCT CTGTGTCATCT	1080
25	GCGTGTGCT GAGAGGGCTC CTGAGCGAGC CGTTTGGGGC AAGAGCTGCC ACTGCGAGG	1140
	TGAGTCAAC AAGCAGTTGT GCTGAGATG GCACTGCTGG AAGAGCTGCT GTGTGCGAGG	1200
30	CTTTGTCTG AATGCTGAC ATGATATGAT TCGTTTACTG CTGACAGAT TGTACCCATTT	1260
	TGCAGAGAA GAGAGAGAG AATATAGTGG CTGTCTAGG GTCATGTGAT TGAATAGTGG	1320
	CAGAGAGAG AACTGAGCA AACCTGTCG TGTGAGAGCC GGTGTGAGG TTTCTTCACT	1380
35	AAGAGTCTCT GATGAGATTA CCGAGAGTGG GGTGCCATCC AACATGAGG TGTGCTTGCA	1440
	TGTATGATCT CGAGCTGCA GGTGTACCT GTGAAAGTTT TGGAGAGCT GCTTTATGAT	1500
40	AAATGAGAA T A	1511
45	(2) INFORMATION FOR SEQ ID NO: 217:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 642 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 217:	
	AAGCTTAATCT TTTCTGCCCA CAAAGATTC GAGAGCAGCG TACATCTGAC TTGCGCACTGT	60
55	GAGAAAGTTC AACTAGAGAG TTGAGCGAAA GACTGGAGCA TTGAGCAGAG GAGAGCGAGC	120
	TGTGATCTGG AACATGAGGC GAGCGCGTTC ACCTTGAGCTG GGTGTGCTCT GATCTTGAAG	180
60	TTTGTGAGG TTGTCTCTGT TTGATGCTT CAGCTAGATG ATTAAGACATG GAGGGGAGTG	240

ACCGCCCTTG GACGTGTTTC TTAACTCTA TCCATATAT ATGGGCGGTGG GATGGTTGTA 300  
GAGTAAAGC AGGATGATCG TGTTTTAAAG CAGAGCTTG GGACGAGGC TCTACACCT 360  
AAATTTCTCT CCGTGTAGCT GAACAAGGT GTAAATTAGC TTACAAAGAG AACAGGCTGC 420  
CGTCAGCCAG AGTTCTGAAG GCCATGCTTT CAGTTTCCCT TGTGACAAAT TCGTCTCCAG 480  
TTCTCATGAA AGCAGAGAGC CTTAGGGGGC CTGGCCAGAG AACACACCA TCTTAGGCTT 540  
GACGTGTGAA CAGCAGGGGG TTGTGTCTCT GTTCTGTGTT TCTGCTTCC GACTTTCTC 600  
ATAAACCTCT ATTTCCTATT TTATATTATG GTNGGTGCTG GG 642

(2) INFORMATION FOR SEQ ID NO: 218:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1241 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

GTGCCACTG TTCCATTTTA TCCATATAGA TTCCATTCTA GGGCCAGCC GTCTCTGAC 60  
TGATGTGTTT CCGTTTACC CTGGCAGCT ATATAGAAAT TTGGGTGAAT GAAGAACCC 120  
AAATAGGCA GATAGTCCCC CCAGGCCCTG ATATCCATPA AGGCTTGGG ATGCAATAT 180  
GTAAATGTCC TTAGTCTTTT TGTGTGTTTA GAAAAAATA ACAGATGGG CTCAGATGA 240  
TGCTTAGCTA AAATGGTTC CTAGCTGTGT ACTCATAACT TTCTTTTGA TTGAGTAGTG 300  
AAAGGAGGA GGAGGAAAGG AATTAAATG TCTTCTAGT ATTCTCTGA CTCAGTCTG 360  
ACATATGCGA TAATTAACCTA TATTGAATG CCAAGAAATG TATCTGAAC AAGGACAG 420  
TTTGACACAT TTATCATGCC TTATATTAC ATATTAACCTG AAACCAATTA ATAAACATAT 480  
GAATATCCA TTGCAAGG CAAGGAGCC TAACTCTTT GTTCTTTTT CTACATAGA 540  
GAATGATTT TTTTUTTTAT TTTTTHAGG GAACTATAT AATTATGNC CAGTGAATGC 600  
TTTTGTGAC TTAAAGCTAT GAATCAGGT TACAATTGAG TTCAATCTAG ATGTTACTA 660  
CCTTGAAAG GATGTGTGTG CTTATGTGA CAGAGGCCAG AGCCTCTCG GAATAAACA 720  
AGCAGATCA TCCACACACC AACTCTAGC TTTAGTGGCA GATGGAGTG GTCACAGCT 780  
CCCAAAATGT GGGGCTTTGG ATTTCCACAC CATCCAGCT GTGTCTATC TTCTCTTTC 840  
ACACTCTTGA TGTAAATTTG AAATGTGTGA AATCACCTCT GAATTTCCCT ATAGCATGAG 900  
CACATCTTGA TGACACATA ACMAATGTT CATATGTGA ATATAGAA CTCCTACAGC 960

CTCAGTTAC CATATATTTT CATGTTTGTG GAATGTATAT TGAATATGCA GGCTTAAGA 1020  
ATTACTGGCA AGTTTATGCC TGTGGTAAAT ACCTTAGGGT TATTTAAATA TTGTATATTT 1080  
5 TATTTAATG TTCAATGATG TTTCGAAGCA ACMAATATAT CAGCGATGCC TCTTTCCCAT 1140  
GGGTCTTATT TTCACTCTCT TTTCTGTAGG AAAAAAGAAC AATGCTCTTAA TGTATTTTAA 1200  
1241  
AAGTTTGTGG TATAGTTTCT AATTCCAATT TTAATAAAG T

(2) INFORMATION FOR SEQ ID NO: 219:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1080 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

TGTTATGTC ACCTAAACAA TACACATG CACACACAA TACATATCCA TTCATCATTT 60  
CATTCAGTG GTGTTTCCAG TGTCTGTGTG TCACTGTGTTA TCGAGTTTCC ATTTCOCAGT 120  
GAATATGAG TCGAGGGCCA CTTTCTTAAC CAGATTGTCT TTTCAGAACA AAGACCKGGG 180  
30 RATTCAGGAA GAGTTTGCAA AGAGGGAGAG CCAAGGAAG AGAGCTTTAA ATTGAAGGT  
TAATTTCTTA AGAGGAACCT GGGCTGAATG ACTACAGTGT TATACCCCTCC AATCTTTGCA 300  
GGTGGGCGTG GNACTGCTCT TGTATCACTC TGTCCAGGT ATAAATCCAT ATATCCACAA 360  
35 AAACACATAT CCATCCATCA ACATATACAT GGTTCGGGAT GAGCAGGTCA ATAGTTTGA  
GAGGAGGTT GTTCTCTTTT TTTTCTCATT ATACTCTTAA ATTGTGTCA GTTATCAAAC 480  
40 AAACAAAGAG AAATATGTT TTGGAAAC CTTCCATAGC CTTTCTTAT CAGTGTCTTT  
AAATATAGA CTAATACAC ACATCCGCC AGTTTTTCT TACAGTGACA GTATCTTAC 600  
45 CTCCATTTA ATATTAGCCT GGTATTTTTT TCACGTATAT TTACCTGTGA CTTGTATTTG  
TTATTTAAG AGGAAAGAAA ACATCCAAA AAGAAAAAT TACGTGAGC GCTTCATAT 720  
ACTATATAT TATATATAT ATGTGACAT TTTCGAATAC TGTGAAGTTT TATCTCTGCG 780  
50 ATATACTTTA TACGGAATTA TTAGCCCTTA AAATATAGAA AATAATTTT ACAAGTTTTC  
TGTTTGTGT GGAAGAGTAA TTGATGTGC TAAGATGAT GTTTGTTTTT TTGGGGTTTT 900  
TGTGTTTTT TTTTAAATG TTACAGCNC TTTTTTTGTA AGTTTCACTT TCCGAGGTAT 960  
55 TGTACAGTT CACACTGTTT GTGAGTTTG AATATGAGG AATAATTAA AAAAAAAA  
AAACCKGGG GGGGCCCGG TCCATTTGAG CCCAAGGGG CGGTTACGGG GTCACGGCGG 1080

(2) INFORMATION FOR SEQ ID NO: 220:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1258 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

10 TGAATGAGG GCTTAAAGAT AACATATGG GATTGAGTT GTGTGTCAT AGGTTTCAC 60  
15 TGCATTTTG ATTGAGTTT ATCCCATTA ATTCTTACA GTGAATTTT ATTAAATAT 120  
20 ATTGACATA TATTTCAGT GGAATTTGCT CTGAAGGTC TCAGTGGTC TGAATACAG 180  
25 ATAGTGGCC TTCACTGTG GGAATATCA GGGGAGGAG GCTTCACCTC TATGACACA 240  
30 TTGTATTATC GGAATGCTC TCCCTGTCT ATTATCTTG AGCTTACCA TCCCATAC 300  
35 TTACACACA GCGAGAGTG GAAACAGAC CTACACACA AGCTACACCT ACCCATACA 360  
40 GAAGCCGTC CTTGCTGCT CTGGCCAC AGTGTGATC TGTCCCTTG GCGAGTAC 420  
45 GCGATSCGA TTGACCGGT CAGTAAAGAG AACGTTTCA CAGTTGAGC AGAAACATCA 480  
50 GTCAAGACA ACAAATATAT TATAGAGCT ATGAGATGC TCATTGAAA GATGATGAG 540  
55 AATTCACAG AAGATATCAT GTCTTTTCC ACCGAGGGG ACTACATCA TCTACAAAC 600  
60 AAGCTTCCA GCTGCTCCG CTGCTATAG TGTTTGGTT ATTTCATC CCAATCTCG 660  
65 GAGGCTTTT AAGTCTTC GCTTGGTGG CCGACCTAC TATTTTATA AATACATTTG 720  
70 AATGTCTCC TGACTACTGT CAGTAAAGA GAGCATGTCT CACTTAAAA AGACACTGG 780  
80 AACCCAGTG CATTTCACA TCTCTGAT TACGCTTTSA CATTGCTG TCTACATTA 840  
85 GTCCCATTA GTCCCTCGG TGTAAATCT TCTATACAG CTTGATTTG TATTCGAAA 900  
90 TTTTGTGAG AAGATKAAA ATCACACCT GCGTTTAGA GATCATATT CTACACTACT 960  
100 TCTGAGTTA TTTTTCAT TGAATTCAT TGAATATAG ACTTCCATT GAGAGAAA 1020  
105 TGAATGAAA TGTCAATTC CAATTTCTT GTAGCGGTT GTTTCAGATT CTCTGTCT 1080  
110 TGAATGTA AACATGAT CTGGAATCA GAAAGAGGG TCTGGGATC TGTGATTTT 1140  
115 TGGCTACAG AAGTCTCCA GAGTCACTG TATTTTGA ACTTTCACG TCAATTTAA 1200  
120 GTTCTCTG TCTTGGCAT CAGATATAT TCCATTTT TGGCCCGGG CAGGCTG 1258

(2) INFORMATION FOR SEQ ID NO: 221:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1693 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

10 CAGATATAT GAATATGAC CCTTAAAA AGAGAAAAA AAAATCCGC GGTCAACT 60  
15 AAGACACAT TGTCTATTA AAGCATATT TATTACATA GAAAAATTT AATTCAGAG 120  
20 ACTATTCAT ACTCATTAC TAGAAGTTC TTTTAAAT GACCTTAAA ACATTCATG 180  
25 AAACTTCAT CCAATACCA CCGTTTAT TTTCTTAAA CATTGTGAA GCTTACCTT 240  
30 CTGAGATCA TGTGGCAT GTATGGGCA GTAAATCAC AGAAGATAG TTTATGACA 300  
35 ATTAAAGCT GTTTCACCT TTAGAGACA GCTGGGCTG AGTATCTC GGGGCCAGAG 360  
40 TGCATTAAG TTTTACAAA ATTAATCAT ATGTCAATG TTTGACGTT TGTTCCTTG 420  
45 TTGAATTTT GAAACGAG TTGACATC ATTAAGATA TTACTTTCT TCAATGGTT 480  
50 TTTGGTTAC TGCCTTAAA GCTTTCAG AGATCTAG GCGACAGAG CATTACATTT 540  
55 TCCATCTTAA TAGAATGAA ATTAAATTTG TATCTATCA TACAGATC TGGGTCAAT 600  
60 GAAAAAAT CATTATTC GCTTTTAA TATATGTTA AAAATATAT TTAATGTCT 660  
65 GGAATATCA GAGACCTCT GAGACACA GTTTCATAT AACTTTCT GACCATATG 720  
70 GCTGGACCG TTGCTGCTC TTTGGAAA GGAAGGCTG TGTAAATAG GCTTACATC 780  
80 TTCAATTAC CAATTTGTA TAGATTAAT AAGATATTA TTTTATGCT CAGTATTTA 840  
85 TTAATTAAT TTTTATGTA TGCATCTC TTGCTTAT TACGATTA GCAATACATA 900  
90 GAGATTCAG GATATTTTG TTTAAATCT TCGAATTC AATTTTTC AGTGGCTGC 960  
100 TTTTGGAAA AGATATCTT CTATACACT TGTTTTACC TTTTGAACA TTGACAGAA 1020  
105 TTATCAATG GTTGTGAG ATACGACT GATGGCTAG TTTAATCAT TTGCTTCAA 1080  
110 AGTGGCTAC TGAAGAGCC GTTACCTAG TGAATTTAA AAGATTTCA AAACTTTCT 1140  
115 ATTCCTTCT TAAACCTAC AGCAACTAG GATGTGATA GCAATGAGT GTATGATGA 1200  
120 GAAATTTGA CCAATTTGT TTTTGTGAG TTGTGTGTA TTTGATTTG AAATCATCT 1260  
130 TATTCCTTT AAGATTTT ATGTATGAT GTGAATAGC TACGACAT ATGCTCAT 1320  
135 ATTCACTGA ACTCTCTT GACCTTAC AAGATTTCAG ATGCTCAT GATGATGT 1380  
140 ATTCTTTAG TAAATATAG TAAATTTAC AATGATCTT TAAATGATG ATGCAATCT 1440  
145 GATTTATAG TGTGTGCT GATCTTAT GAGGCATAT AAACAGTTT CATATATAT 1500  
150 TTTCAAGCT TGAATTCAC AACTTACT TTGAAAATTT CTTCCATCC TGAATACGA 1560  
155 ATTAAGAGG CCAATATAT TGCCTCTTA TCTTTAGAT TTCACTACT TTAATTTAA 1620

449

AGTTGTGTAT AATTGTATAA ATCTGTGAAA GAATAAAG TCGATTATAA TTAAAAAAA

1680

AAAAAAAA AAA

1693

5

(2) INFORMATION FOR SEQ ID NO: 222:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1196 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

ACCGTGGGT CGACCCAGGC GTCCGCGAG TGGCTGTGT GGAAGGGAG AAGATTGT

60

AAACCCCGGA CGAGGTCTT GTTACCGGA GCGCGCTGCT GTGCGAGAC CCGCGGGTGA

120

AGCCACGTC ATCATGTCTG ACCAGGAGC AAAACCTTCA ACTGAGACT TGGGGATTA

180

GAAGCAAGT GAATATATTA AACTCAAGT CATTCGACG GATGAGTGT AGATTACTT

240

CAAGTGAATA ATGACACAC ATCTCAAGAA ACTCAAGAA TCATTACTG AAAGACAGGG

300

TGTTCGAATG AATTACTCA GGTTCCTCT TCGAGGTGAG AGATTGCTG ATATCATAC

360

TCCAAAGAAA CTGGGAATGG AGGAGAGAA TGTGATTGAA GTTTATCAG AACAAAGGG

420

GGGTCAATCA ACAGTTTAGA TAATCTTTT ATTCTTTTC TTTTCCCTCA ATCTTTTTC

480

ATTTTAAAA ATAGTCTTTC TGTAAATGG TGTTCAAAC GGAATTGAA ACTGCGACC

540

CATCTCTTC AAACATCTGG TAATTGAA TCTAGTCTC ATTATTCATT ATTGTTGTT

600

TTCATGTGC TGAATTTTGG TCAATCAAGC TCAGTCCCT TCAATTATCC CTCCTCTTC

660

TAAATATAC GTGTGCACAG AGAGTCCACC TTTTTCAGGA CATTCGATT TCAGCTTGT

720

GTTGATAAAT AAGATCGACC AATCGAAGT TTCATATGA CTTTCATT GGCCTGATG

780

TTCTAGCATG TGAATTAATC ACTCTCGAC TGTGACTTC AATGCGAGAT GGAATTTTC

840

CAGAGACTG AACTGTGAAA AATGACTCT TCTTAACCT GAAGTACTT TTAATTTTG

900

AGGCTCTGGA CCAAAAGAG AGGAATATCA GTTCGAAGT AAGATGACG ATAGGTGAG

960

AGTAATGACT AACTCCAAAG ATGCTCTAC TGAAGAAAG GCATTTTAAG ATTTTAAAA

1020

AATCTTGCA GAGATCCCA GAAAGTCTT AATTTCAAT AGCAATTAAT AAAGCTATAC

1080

ATCCAGAAAT GATACACCA GAACACTCT CTTTTCAT TTAATTGAC TTTTGGCTT

1140

GGGATATGG TTTTAATGG ACATGTCTG TACGAGCTC ATTAATATTA ACATA

1196

60

450

(2) INFORMATION FOR SEQ ID NO: 223:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1791 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

TAGGGAGGT GGCAGGAAG GCTTGGACA GCTGCGGAG TGAAGGAGC GCGGCCCCG

60

CCGGTTGGC TCGAGTCTGA AGCTTCCAG TAGCGGCGG CAGAGCTTGA CCCAGGCTCT

120

GGCATCTGT AGCCCAAGTC CCCCACATC AGTCCAGTGA TGAATGCGGA AGTGAAGGTG

180

ACAGGCGAGA ACCAGGAGCA ATTTCTGCTC CTAGCCAACT CGGCCAAGGG GGCAGGCGTG

240

GCCACACTCA TCCATCAGT GCTCGAGGCC CCTGTGTCT AGCTGTTTGG AGAATGCTG

300

GACATGCCCA ATGTATGAGA GCTGGCTGAG AGTGACTTTG CTTCTACTTT CCGGCTGCTC

360

ACATGTTTG CTATGCGGC ATAGGCTGAC TACTTAGCTG AAGCCCGGAA TCTTCTCTCA

420

CTAACAGAG CTACAGAAA TAAGCTTGA CACCTCTCAG TTGTCACTCT GGCCTCTAAA

480

GTAAAGTGA TCCCATATGC AGTGTGCTG GAGGTCTTGC CTTGGTAAAT GTGCGCAGC

540

TGGAGAGCT TGTGATTGAG GCTGTGTATG CTGAGTGTCT TGTGTGCTCC CTGAGCAGC

600

GCAACAGCG GCTCGAGGTT GACTACAGCA TCGGCGCGGA CATCCAGGC CAGGACCTCA

660

GTGCCATTC CCGAACCTTG CAGGAATGCT GTGTGGGCTG TTAGGTCTTG CTGTCAAGCA

720

TTGAGAGCA GTTGAGCGCT GCGAACGAC ACGAGAGCA GCGAGTGGGC CTGAGAGGCG

780

AGATTGAGG TGAAGTCCC AACCTTAAA AAACATTAA ACTTAAGCG GCAAGCAGAG

840

CCGAGCGAC ATCTCAGGAC CCTGAGCAAC ACCTGACTGA GCTGAGGGAA CCAAGCTCTG

900

GCAACAGCA GCGCCAGCCA CGAGAAAGC CTCAAAAGGC AAGGGGCTCC GAGGGAGGCG

960

CAGATTGCG TCCAGTCTGA ATTGAAGGA CTGTCTTTTC CTCCTGCGG ATGTGGGCTC

1020

CAGCTGCTCT GCTGCTCTCT TAGGAGTCTT CAGAGAGCT TCTGTGCTCC TGGCCAGCTG

1080

ATATCTTAG GTTCATGACC CTTCACCTCC CTTAACCCCA AACATAGATC ACACCTTCTC

1140

TAGGGAGGAG KCAATGTAG GTCACTTTTT TGTGTGACT TTTCTGTTTT TGTGACTTCA

1200

TGTGTTCAAT TGTCTGCGGC TGCATGCTC TCTGCTTGT TTTCTTAAGA GCTCAGCATC

1260

TGTGCTCTTT CATTAATGT CATTTAGTAG GTGGTAGCC CTGATGCGGG TCGCTCTGTC

1320

TGGAGCATTA CCCACAGGCG TTTTCTCTGC CAGCCCATCC CTGCATGCTT GATCCCGAGT

1380

TCTTATACC TACCCCTGAC CTATTGAGCA GCTCTGAG ACCCTAGGG CCCCAGCTTT

1440

TACTCACAC CTGAGATTC TGGAGCGAG TCTGCCATGC CAGGAGTCAC TGGACATGTT

1500

CAATCTAGAA TCCCTGACA CTACAGTCAT TTTCTTTCT CTCTCTGGCC CTTCGGCTCT  
1560  
GGAAATCTCT CTCTCTGAC CCCAGAGCT AAGAAATGCA GCGCTTTCT AACATTTGA  
1620  
GAGATGATTC TTTCTTGGCC CTGGCCAGCT CGGAGAGCTT GATGGCAATC CTGGAGAGCT  
1680  
TTAATCTCTCT TTTGTGAGTT TGCTGGGAAA GGGAAAGGTA TATGATTCAT ATTAAAAA  
1740  
AAAAAGTATA TATGATATA TCTATATATA AATGAGCCA GAAATTAATC T  
1791

15 (2) INFORMATION FOR SEQ ID NO: 224:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2517 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

25 AACTATGGA ATCCAAAGAA TTGGGACAG CGGCAACGA TTGTGAGCT TTTCTGTGTC  
60  
TTGGGGGCC TCAAGGAGC TGGACTGTTC CATCTGGGG TAGCAAGTTC GTTGTGTGA  
120  
AAAAAGACGA TGGCATTCAC GCAATATAGA AGTAAACAAA CCGATGTGTC GACGGGACAC  
180  
CGATGAGTC CAACCTTCAC ATGAAATGGA ATGTATTCAC CTGACACGAG CCCATCTTC  
240  
TCCGACGTAG TGAAGAGCCA TGAATGAAA AGAGAGACCA GCTGCCTTCC AGGTGAACT  
300  
CTGCCTCTTC CTCCAAAGCC CCGTCCGAG TGGAGCTTGA CACATCTTC AAGGACACTT  
360  
TGAAGTCTC AAGGAGCTCT KTGACACATC AGCCCAACGA ATTGAAATTC AACCTTTGAG  
420  
CAAGGAGATC AAGGAGCCAG AAGTGGGGGC AAGAGAGGCT GACTGTGTT CCCCAAGCA  
480  
AAGCTTATGA CCAATGGGCC ATTCAGATGAG AAGACCTTGA TTGTGGGAG GTTTCGACAG  
540  
GATTAAGAGC TTCTCTACATC GATGGAGCC GCTTTCTGT GTTGTGTCT GCGCTGTACT  
600  
CTTCTCTCTA GATTAAAGTT TCCCTATGTA TGTTTCTCA TCTCAATGCC AAGTAAAGCT  
660  
TGTGTTTTAA AGTGTGACC TCCCAAGCC TGAAGCCCAA GCTGATTTCT TATCTGAAA  
720  
TGTGACATG AATTCATGAG GTGGCTTTCT TGTGAGCCCA TGGATGACAG CTGGGAGGCT  
780  
GTCTGAGAGA CCGTGCCTTT TCCAGGGGCC GAGGGGCTTC CTTTCCTTTC TGTATTTTAA  
840  
GCTTTTAAA CAATGAGAGG GATGAGAGC CCGTGTCTCC TGAAGGAGCC CAGCTTGGCC  
900  
TGAAGACTCT GCGGCTCTCTC TGTCTGTCA GTGAGAGTCC CTGGGTGGGA AGCAAGTCTC  
960  
AAGGCTCTTC TCCCTCTCCC AGTGGCTTCA GAGCTTACTA GTGGCAGAG CAGAGTGAAG  
1020  
CTGACCCGCT GGAAGAGATC TATCTAAGCT CTGGCTTTCG AGTCCCTGT CTCTCTCAC  
1080

CAGAGAGAT TCTGAGATC TGAATTTCC CTTTCTCTTC AATTGTGTC AATGCCCCAC  
1140  
CCAGCTCTTC TTTCCTCTTC GAGTGTCTTT GCTGGAGAG GCTGTGTTC TGAAGCTTCC  
1200  
CGATTCTAC CTGGCTTCAG ACTTAACAC AACCTGTTT TGTATGAGG CCAAGCTCTCT  
1260  
TCTGTGTGAG CATTGAAAG GCTGAGCTTC CCAATGTCA GTGCTGGAT TTGAGACTTA  
1320  
TTGATATGA AAGATGATC TTGTCTCTG CTCTCAATT CTGGCTTAC TTGTCTACAG  
1380  
GAGATGTCT AAGATTCCT GAGAGCAAT ATCCAGCTTC GACCAAGGG CACTGTGCT  
1440  
TCCCACTTAT GTGATGAGC CCAATCAGCC ATGACAGAG GATTATTTTC CTGCTTTCAC  
1500  
AAGAGAGAG GATTCAGAG AGCAAGGAGG CTCTACAGC CAGGTGTTT CCCCAAGTA  
1560  
GCGCAAGCA GTTGGAGACA AACTTCAGAG CCCAGGAGAT CCGTGAATGA CCAAGGCAAT  
1620  
GTTCCTACAT AGTGTCTCCC TCGTGTTCG GACTGAGAG AATCCAGCT GGCAGAGCTC  
1680  
GAGCCAGTTC GGAAGACAG TTCTGGAGC TCTGCAAAAT CAGTACAG TCTGTGAAA  
1740  
GCAATATCC GAAATATATC AAGAGTCCC AAGATTTGCT TGAAGCTTAC CCAATGAAA  
1800  
AATCCAGACA CTCAATGTTT CAGGGTTCAG TCTGTACAG AGAGAGAGC CAGATTTTGA  
1860  
ACCGAGCTC AGTGTGAGG CTCTGAGCT GCGCAGAGG GAAATCTCA AAGAAAGGAC  
1920  
CTGGGTGTC TCCACTTGA GTTCTTTAA GATGTGTCT TTTTATCTC CTAAAGCTTT  
1980  
CAAGTGGTC CAGACTTCT GTTACAGCT GGAAGCAAT CTTCCACAC TTTTCCCTTC  
2040  
CTGGCCAGG AAGATTCGA GAAAGCAAT GAACTGTCT TTTCAGATC TGGAGAGCCG  
2100  
GGCTTACATC CTTCAGATCT GCTTAAGGTA GGAATGTTA AATATCTCC TCCATGGCTT  
2160  
TATCTCTCT CTCAATCCAA AGCAAGTATC TTCTGTGTT CAGAGATTT CATTGAGTCC  
2220  
AAGTCAAGC AGTGTGCAAT CTGAGCTGAG TGTATAGCT GACATCTGAG TACATTTGAG  
2280  
GAACTGTCT GCTCTCTCA CTCTAAGC AGTCAATCTG GAGAGCCGG AAGAGAGGCT  
2340  
GATGGCTAG TCCCTGTCTC TCTCTCACTT CCAATGCTTC TATGTAAACC ATCTGTGCT  
2400  
CCATGTGAGA AAGAGAGAA GGGGCAATTA GAGATGAGG GTGATTTGCT ATTACTATC  
2460  
CAATCTGA TTAACATTTG TTATCTTAA AAAAAAAAA AAAAAATCTG AAGGGGCG  
2517

50 (2) INFORMATION FOR SEQ ID NO: 225:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2424 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

60 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

5 TTTCTACTTAA TCGAGGATG ATTCTAATGA CAGAGTCTTT CAACACTTGG CACTGATGT 60  
ATCAGCAAGC TACAGCTTGC CATGTGACTG GAGATTACT AGACTTCTG TCATATTTC 120  
TTTGGTTT GAAGTCTACA CCGCTTATC TTAGAGAA AGATGTGAA CAAGCATTA 180  
TCCAGTGGCA GAGAGGAA GTTGTGCG ATTAAGTGT ACTCTTCTT AATTCCTATA 240  
GTCTCCAGA ACTTGAAT GCTGTATAG ATGTCTCAA GGAATGTGA CTTTGAGTC 300  
CCATGATT TTTTCACT ACTGTTCCT TTCTACAACA CAACATGT ACTTACCATC 360  
ACAGTAAAT ACCAATGTCT CTGGAATTT ATTTCCCTTG TCRGAAAT ATCAGCTAA 420  
TAGAGGGA AGCAATATT CCGCTCGC CCGCTGACT CATATGTGC CTCTTCCCA 480  
CAATGTGGA AACAGTAAG GGCAGATG ACCTTTATGA TGTATCTG CTAGACTACT 540  
TCTTTCTTA TCATGATTC ATCATCTAT TATCCGAGT TCCATTCAC TGTGAAAT 600  
TTACTGAAC ATTAGTTAG CTGATGTGC TATGTCCCTA TGAAGTTTG CACTTCTC 660  
TTCACTGTT CCCAACTT TCGATGAGC TATCCAGAC TCACTCTCT ATGTCAAAA 720  
ACTGATCAA GCTTTTGT GAGATCTCT TTTTGGAGA ATATATATA TGTATCTTAA 780  
TGAATGAAG ACTTTTATA AACAGACA TTGTCTACAC GTTCATGACA CATTTCTTC 840  
TAAAGTTCA AGTCAATG TTTTCTGAG CAACCTGTGC CATTTGATC AGCACTCTA 900  
TTACAACTT GATAGGCG TATCAGAAC TACAGTCTGA TTTTCCAAC CGAGTGGAA 960  
TTTCCAAAG AGTCTCTCT TTAATGGGG ACTGAGGGC ACTGCTTTC CTCTGTGAG 1020  
TACACTCC CAACAGTTA ACCGAGTC TATTTCCAC TCTGAGAG CTTTTAAACA 1080  
AATCGAGAC TTGTCTGGA CAGAGAACT CACTCCAGA GCAAGAGCC AAGAGAGAA 1140  
AACTAAAGA TATGAGGA GCACTCCCA TTAAGAGGG CCGTGTAGC AGTCATGAG 1200  
ACCACTGT AGACAGTGC ATCATGACA TGAAGAGCA AACCGGAG GTCTGAGCC 1260  
CAAGGAGC TTCTGACAT GAGAGCAGAG ACTCTCAAT TATTATCCA GGAATGAGC 1320  
AGATCTCC TTCCCTGAA ATAGTTCTG TTAAGAAATA CCGAATGAA GTTCAATCT 1380  
CCTTTTACA AGCATGTCA ATATAGGT CACAGCTGC AGAAGAACG TCCAACATG 1440  
GTAGATATGA CGATTGTAA GAATTAAAG ACCTCCACTG TTCCAGGAT TCTACCTAG 1500  
CCGAGGAGA ATCTGATTC CTTTCACTT CTATCTCTG ACTTCTGCT GACTTACTG 1560  
ACTTGAAG CTGTATGG CAGCTTTC CTTCCAGGA CCGTGAAGT GCTTTATTC 1620  
TCAATGTGG CCATTCCAGA GCACTTTTA GTCATATCA GCACATGAC ATTTTATGA 1680  
CCCTGTATG GACCATGAA TCTACATCC ATGTCTCAC AAGGATATC TGGCAAGGA 1740  
AACCAAGCT CTCTTGACA TTAGGTAG CATGTCTACT TTTAAGTCC TCACCCCAA 1800

CCCCATGCT GTTGTATAA GTTTGCTTA TTTGTTTTG TCTTTCAGT TCTCCAGTGC 1860  
TCTCTCTTG AATGCCAGA TAGATTTATA GCGTTAATTC TTGTCAGGC AGAATCTCAG 1920  
ATGAAAAAA CTGTGATCTT CAGTACTT CTTAAGGGC AATCAGTAA TGGATATGTT 1980  
TTATGTAAT AAGAGTTCAC TTTAGTGGT TTCATTTAAT ATGCTGTCT GGGAGAGACA 2040  
GGTTGCTTA GCGCTGACA ATGTAAATTA AACTTACAGC ATTTTACTG TGTATGATAT 2100  
GGTGTCTCT GTCCAGTTT TGTACTTAT AGAGGAGAT TGCCTCCAT CCGTGTGTT 2160  
CTTATATCA AATTAAGTT TACTTGTATA CGGACACACC ACAAGAAAT TGAATCTGTA 2220  
AAGANTCTC TTAGCTGTG GCTGGCACT ATATAATGG TCGTTTATTT AACAGATAC 2280  
CTGTGGAGA AATAAGCAC ACTGTATGTA AAATTAATTT TTTTATTTTT ATTGACATGA 2340  
CTGATGATT GCTATCTGT GCACTAATTT AAATGATTT TGAATGACTTA AAAAAAATA 2400  
AAAAAATA AAAAAAATA AAAA 2424

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1080 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 226:  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

5 ATATAGAGAG GATATCTGT TTAATTTCTG TTCTTCTGA TGCATCTACA AGCGGTAAAC 60  
TAGGTGACA GAAACAAAG ATCTTATCCA AAAGAGTCT TACAGCAACC CAAGTCTCA 120  
TCTTCCATA GTAAAGTGA CCGCGCTTG AGGTAACTA CAGGCACAC CACTTCCGG 180  
TTTCTCTTC CCGCTGTCC AAGATGGGG ATGAAGCCAC GCGAGTGT TGTCTGTAGA 240  
TCCCGTCT GAAGCTAAC CCGGAGCCC GAGATCTGTA GTTGTGGTG CAGGACTGA 300  
AGGAGGATA TCACTCCCTT ATCCGTATG TGGAGAACAA CAAGAATGCT GACAACGAT 360  
GGTTCGACT GGAATCTAAC AAGAGGAA CTGCGTGT TGGAAATGC TGGTATATCC 420  
ATGACCTCT GAATATGAG TTTGACATCG AGTTTGACAT TCTATCTACA TATCTCTATA 480  
CTGCCCCA AATTGCAATT CCTGAGCTGG ATCGAAGAC AGCAAGATG TACAGGGTG 540  
GCAAAATAG CCTGACGAT CATTTCAAC CTTTGTGGG CAGGAATGT CCCAAATTTG 600  
GACTAGCTCA TCTCTGCT CTGGGGTGG GTCCATGCT GGCAGTGGAA ATCTCTGATC 660  
TGATTCAGA GCGGTCAT CAACACAAAG AAGAAATCAA CCAATGAAGA ATCAAGCCAC 720



780 TGAAGCAGGG CAGAGGAGC TTGATTAAGC TTAGATTAATA TTTTCTCTTC CATACACATT  
840 AACTCATCTA ACTCTCTCC CGGACACCT CCACCTTAG TTTTACTTA GTACCTGACG  
900 TACGATTCGC TGGGAAATA ACAACACAC ACCAAACAT ACTCTACTT AGTTCTTAG  
960 GTTCACAGG GAGGAGAGG ACTGCGCTT GAGCATCTA GAGGTAAATT ATATCCGCC  
1020 CCAAGTCAG CAAATGCGA TTCTGAGGC ACCGGGTTA CTGAAATGA GTACATTAAG  
1080 TCTTCTGCT TTCTGAGAT AACCATGCA TAAACCTGC TTCTCTGCG TAAAAAAAG

15 (2) INFORMATION FOR SEQ ID NO: 227:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1316 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

60 TTGCATCAC AATTACTGCG AGGCAAGCAG GCGCAATTGC ATCTGCGGG TGGCTGCTAG  
120 GCTCCGACGC TCCTCTGCGT TTGAGGATG CGGTACACT GCTTCACATG ACCAAAGATG  
180 TTATCCCTGA TCTCTGAGGT TGAATCTGC AGGCGAGAC TCCACATATC AAAAGATCT  
240 GTAAATTCAC CAGAAACAT CTCTGATATC ATCCAGGCA CTACTGAAAT GACTGAAAC  
300 AAGATGATTT TCTTTTCCAA TGTTTTCCAA TCTGGCAATT CTTCGCCAAA GCATTAAATAG  
360 ATTCGAACG GTGCTGCTT CTCTATCTGT CCTTCTGCT GGGCAATGAG ATCCGTAAAG  
420 AATGTTTCCA GACAAATAG CTTCACCTTC TTTGCTCTCT CAATCAAGTT GGAACAAACA  
480 AGTGAATGCG CAGATGCCC AGACAGTAC ACCTTGCAT GGCACAGCTT GATGCAATTA  
540 ATGGCATAC CGGTACATC CAGATCATG CTCTCTGCA TGAATCTCAG CAGCTTCTCA  
600 GTGACACAT TCTGCTCTTC ATTGATTAATA TGTCTAGAC CTGGAAATTT GACTTCCTCC  
660 AAGCTACCG GACCAAAAGG CTCTCTCTCG TGAAGCATG GACCCAGGT CCCTATGAG  
720 AGTGGCAGC CCTGAGGTT GCTACAGAT ATGATCTGC CGTACTCTT TCCACAGGTA  
780 CTGAAACTTG ATGTCCAGGT CAGTCAATGG GAGAGACTG ATCCAGATTT CTGAGACAT  
840 AATGAGAGAC TGTATTAATG CTGCGGAGAC TTGCATCTCC AGGGTTTCAG TTTTGGGACA  
900 CACTGCTTCC GAGTGCATG TCCCACTCT GCAATTCGCC ACATGCTCG GCGCAATGCG  
960 AAGAACATG ATGTTCAGAA AGGGAGAGT GTCTGAGAG GAAACATGTT GCTCCACATG  
1020 ATCCAGTCA TCTTCTCAT CTCTTCAAT CAGATCAATTA TCTCTCAT CCGAGGAGAC  
1080 AAGCCTCTG GATCTGCTGT TAAATATCA GCGCTGCGGC TGAAGGATGT CACACACTTG

1140 AATATCTTC ACTGGTCTCA TGGGCACTTC CCTGTGCGC ATCCATGAT CAGATTTGAA  
1200 TTTCTGCTC TTAATGAGAG CAGACGCGAG CTGGGCTTTC CATTTACTG GCTTCAGGCTC  
1260 ATCCAGCCCT TCTGTACTT TCCCTCTTC TACAGCCGAG GCGTTAAAA TGGTAAATTTT  
1320 CTCTCTTGT TGAAGCAAT GCGGCTGCGC ATGTTCAGG GAAATCTGGA AGCTTTTGA  
1336 GTCCCTGCTT ACCGAG

15 (2) INFORMATION FOR SEQ ID NO: 228:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2043 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

60 TGAAGTGTG CCTCTCTGT GTCTGAGGG AGCTGTGCG GACTCTTTC TGGAGCAAT  
120 GACTTCAGAC CCGACCCACA CTTCAGATG AGACCTCTGC CTGCCCCCGG CAATATCTCT  
180 CCGAGTCCCT TCGAGCTGC ACTTTCACA TGTCTACCC CAGACATATC CCAATGCGCC  
240 CTGACATCCC CTTCCTTAGG CTCTTTTCCA AAGATCTGAG GTACATGCTT GCTGTGCAAT  
300 CAGAGTCCCA GGTTCACAGA GCTCCGCGGG AACGGTGTCT GCGTGTCTCT CCAATTAAGT  
360 CAGATTCAGG TCTGAGACCC GTGTGATGA AAGTCTGAG CAGCAGCCTT GCGCTCTGCC  
420 CAGGCTGCGG TCTTGAAGAG CTGGTTTTC TCCAGAGAG TTAGAGCAG CAAATCTCTT  
480 CTCTGCGCTC AAGGTGACAT GCTCCAGATG GGGCGATGC AAGAGAGCAG GTGGGCTGTG  
540 GCTGACATG TCCGAACTG GCTTCTTAC CAGAAAAGCC TCAAGCTTCC TCTGAAACA  
600 TCCCGCTTTC TGGCAGAGG GAAAGGCTC CTTTAAGGG TGTCTTTC CAGTGGGAG  
660 CAGTCTGCC CTGCCCCCTA CTAAAGCTTC TGTCTGACG ACTTTCCCC AAGTCTCTGT  
720 AACTCTTGG AAGGTGAGTT TGGCTGCGA GCAAGTCTCT GACAAACTC TCCCTGCCCT  
780 TTTAAATTC ACTCATTTTG TAAAAACCA GCAAGCTGAT GTTACTTTAG CCGTCAACT  
840 TTTTCAATTT TTTCTTTTCC TCTTCTCTGT TGAATGAGG GTTCAATATT GCTCTCTGCC  
900 CCTGATGAG GAGATGCTCT CTTTCTGCG CCACTGCGG GCTGTGCTCA GCAAGCTGCG  
960 TCCCAAGCTG GGGGACCGGG ATGGGGGCTT CTCTCTCTG GAGGGTTCGA GATGCGCTCC  
1020 CCAAGCTGCG AAGGTCTCTT CCGTACCTCC CCAATCTGCC CCGCTGATGA GATGTGAGCT  
1080 TCTTGTCTTT GGAATCTCTT GCAATGCGA AAGAGACAT TCCAGATTT GTTGTCTTGG

TTTTACTCAC CTAAACCTTTA GAAATATGAT GTTAGAGAGT GCTTGGCAG CCGGACACAGA 1140  
GTGTTTGCTC GCGCTGAGAG AGGCTCTGCT CAGCCCTGAG AGTCCCTTCC TGCCGACAGC 1200  
5 ATACTGGCAC TTTAAAGAG AGCTGACCG CACAGTGTCC AGACGAMTTC GCGCCGACGA 1260  
GATGGGAGT TCTGTCTGCG CTTTCTGTGT CTGCGTGAAC TCACCGAGCC TAGAGGGGAG 1320  
GTGCATTGAG GTTAGAGTTG CTTCTCATTC AAGTTCTGG GCTTTGGGT GGAAGACAGC 1380  
CAGCTTTGGC GCTGTTGGGG AGACTCTCC AGACCGAGAA CCGCAGAGG AGACAGAGCC 1440  
TGCCACATCC TCCGAGCCCA GCGCTGGGCG CAGGCTGATT GCACTGAGAA TTTGGCCACA 1500  
15 ACCAANTTGA TCGTGGCTGG AACGAGGCG CAGAAAGCCT GCGCTTGTCC CAACTGTGGG 1560  
GCGCTGTCT CAGCCCTCTT GTCCCTTGCA GCTCACTGAA TTCCACACAG GTGCCACAG 1620  
CTCTGGACT TCAAATTTCTA TATATTGAGA GAGTTGGAGA GTATATCAGA GATATTTTTC 1680  
20 GAAAGGATT GGTCTATGCA ATGCTAGTTT GGAATCTTCT TGAAGTTTAA ATGTTTAT 1740  
TAGGAGATT AAGAAGATA AAGCTACAA ATATCTTTAG GTTTTTTTTT TTTCTGTTTT 1800  
ACCGACAAA CTGACACANT GGCATCTCTA TCAGCATGGA GGTGTCTCAT GTTCTCTCT 1860  
25 GTCTTTAGG AGGTATAG GAGATGSGG RAGGGGTGTT TTTTCTTTTG ACTGCCCTCC 1920  
TTTCTAACAG AATGTTGCCA CCACTCTTG AGTGGCTGT GTTGTCTCT CTGTCCGAC 1980  
30 TTCTTTGTA GAAATATCA TTGTTAGGG AACTCAGGCT AGTGTACGG TCTTGTTTTG 2040  
GGG 2043

## (2) INFORMATION FOR SEQ ID NO: 229:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 540 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

TAAAGAGAG CCGAGAGATC TGCGCTGCG TCTAGAGATC GATGGCTAG AGGAGAGCT 60  
50 GTCCCACTGT CCGAGAGAGC TGAGAGCGGT GAATCTCAGA CTCACAGCC GGGAGCTGAG 120  
CCGAGAGCC AGGAGGTCCC TGAGAGAGA GAAAGACAG CTATGACAA AAGCTCCAA 180  
CTAGAGAGG GAATGAAGT TTCTTGCGCA AGAGAGACCG AAGAATCCG TCTCTCTCT 240  
55 GGCATCTT ATCTCTCTGA GCTGTCTTA TGCTTACTGG ACCATGAG CTTGCACTT 300  
CCCGACAGC AGCAGAGGCT TCACCTGGC CCGTTGGTCA GATCAGCA GGCATTTCAA 360  
60 GCGTCAMTAG GACCAAGTG CTGGGCTGTT CCGCTCCCAA CTTAGTGTTC AAGCATGCT 420

TCCTGGGGGC CAGGCGCTTG CTTCCCTGCG CTGCTGGGGG GTTCCGGGTC TTCAGAGAGA 480  
CATGTGCTGT GTCCCTGCTT TAGCCGAGG GAGAGGCAAT AAGACACAA AGCTGGAAAT 540  
5  
10 (2) INFORMATION FOR SEQ ID NO: 230:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 448 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:  
AATTTGAAA TATTAGATA TTSTTACTAT TTGACCCGAC TCAGAAATCT CATGGAGAAA 60  
20 TACTGTGGA TACCCAGCT ATTGTGAAA ATATCAGAT GCAGTATCAC AGCTGTGTA 120  
GACTCTAGTA CCACTGGCC ATCAAGCCA CAGCTAAAA TTGAAACAA AGATCTGAC 180  
25 AACAAACAG CCAAGGTGG GGTGAGAA GCTCTGAGCT GTACTAGCT GTAGAAATCT 240  
ATGCAAGCT GCCAGTGTG GTGTGATAT CAGGAGAAA CTCGAGAGG CCGCGTCTT 300  
30 CACTCTGCT TGACCATGAG CTCTGTGTA GCAAGAGTG AAGGCTAAG CAGATTAG 360  
CTCTGAAGC ATTCGACAC ATACACAAA ATCTGTCAA GCATTAGGA AATCTGTGTA 420  
CTGCTAGTG TTGCTGACC AGAAGCA 448

## (2) INFORMATION FOR SEQ ID NO: 231:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 407 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

GTATGCTGCC CCAAGCAAT ATGTGTGGCT GCCTTACCC TGACTTCTCC AACATGTAGC 60  
50 CCGAGAGGA GCGCTGTAGA CTRAGGGAGG GCGTGTGTGAC CAGGTGTGG TGGGGCTGCA 120  
TGAACTACC AGAGAGACAG ACATTTCTGA ACTCACCTG GCGGATCCAG TCGATCTGCC 180  
240 TATGGTCTGG TCCAGCCGAG ACCTGTGAGA TTTTCTCTAT GAGGATGAC TTGTGCTTCT 300  
CGAAGTATG CTGCAAGCTT ATAGTACTC CCACACGAC CAGCAATAGA GTTAGCTACC 360  
TOTGGCCTTG GATCTAGCC AGCATGGCTG GAGAGAGGAG CAGCTGGCA TGTACCTTAA 407  
60 ATGCTCTTAC CAGGAGAGGA CTCGAGAGT GAAGCAAGT AGCGACT

## 5 (2) INFORMATION FOR SEQ ID NO: 232:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 830 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

15 GTATTGATT TCAAGCTGCT AATGGGCTC ATTTCAGATT CATTCCCTGA TGTAGACATT 60  
AAAAAAAAA CTGAATACGA TTCTTTCAG GATTAACGATT AACACAGCA TGTCAAGCTT 120  
ATTAAATGAT CAGACATGCA GAGACGTTT AAGGTGCCA CAGACAGGA TCACACATA 180  
20 GAGACACTG TACCTCGCTC TGTCCAGAG ACCGAGACA TGTATCGATT TTTGTGTGTG 240  
GTTTATGATT TTTCTGTGTA AAAATGTGTA AAGTTTGTG TTACGTAGT GATATTTTAA 300  
25 TACCTCGAG TGCCTTGGA CTAAATGAT GTCCATGCTT AACACAGTA CATTATGTTT 360  
TGTTTGTTT TGTTTTACAC TCAATATAAA TCAAGAGAG TTACGCAACC ATCTAGCATT 420  
TGAATTCCTC TTTTATATG TCTTCAAGG ATATGAGATG TCCATATCA GAGACAAAC 480  
30 AACACAAA AACTTTGATA ATATATGCTT GATAGACTGT AAGCACTGC TTAACTTGT 540  
GTTCGAAATA TTTAGTGTG ATATATGAT ATATATACAC ACACACAC ATATATATTC 600  
35 AACAAATTA GCAAAATATA ACATGCAATT CACATTTGT CTTTCCCTGT TACGATTTTA 660  
ATACGACAC TGTATACAA GTTTAGTGA TGTACAGATA TGTAAATTC AATTAATGT 720  
AAAACGATT AACACAGCA AAGAACTGT CATTATGAT GATGTCATGC TTTCTATATT 780  
40 AATTAAGTGA CTTCGTTAT CGATCAATG CAGACTTATA CTGTATCTG 830

## 45 (2) INFORMATION FOR SEQ ID NO: 233:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 932 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

55 CGAAGAGAA GACCAATCTA GAATATGAA CTCTATACAC TTCTAGTATT TCAACTTCTT 60  
AAGCAAAATG AACTTGCCCT TACACTGAG GATATACAA TGTTCCTTAT GTACCAATG 120  
60 CTACGAAAC AAAAGAGTG AAGAGACCC TTTTATTATA CTTAATGAC ATATATGAC 180

TTTTAGACA AAGATGCCAG AATAGCCTT CATTCTTACC CTGCAAAAATA ATCCAGATT 240

5 GCTTCTTAAA ATGATATGAG TTTCTAAAGT GAAAGATGCA AATATATGC TGTCACTGAC 300

TCTGAAATG GAGAGGAGG TACTCTGTG TTACGAAAC TGTATTTGTA TATATGTCAG 360

10 GCTGTGATT GTACATATCA GATTTGTGT GCAATATGAC TTTTCTTCAAT CATGCACTGT 420

GCAAAATGA TACTTTTAA GCAATTTCTT CTATAGGAC AGGTGCTGCT ACCTGAAACTT 480

AATTGACAA AGGTGTGAT ATGCTTTCTT AACCTGAAAT GTATTACAT TCAAGAGCC 540

15 TACATTTCT CATTAAGGTT TGTATGCTCA GTATCTTTC AATGACAGG CAGAGCTTAC 600

CTTTCTGAT GAAACATACC ATTTTGTGTA TTTGACAACT ATAGACATGC ACTTGTGAG 660

20 TCCGATTTA AAAATGACA ACTGCTTAT CCAAGATGC TGAATATAC TGTCTATGCC 720

AGTTTCTTA AACTATAAA GCAATTTTG CTTTGTGTTA TTAAATGAG GATAGGCCAA 780

GATTTGGA TTAAGCTGAG GCTTAAATC AATATGATG CTGTATGAT GACGACTGTC 840

25 TCAATATCA GAGCTCGAG AGCAATTTT CAGACTGTG ATTGCCAGA ACACATAGTC 900

CCGAGTTTC TAAATGAG CAAATCTAAA AG 932

## 30 (2) INFORMATION FOR SEQ ID NO: 234:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2786 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

40 TTACAGAGGT GAGCTGTGTA AACACACAC ATCTGTATC CCTCTTCTT TTATTCCTCC 60  
CTGGTTTCA GAAAGAGAG ATATATGAG AACACTGCC CTTCTTTGA TCCGAGATC 120  
45 TCAATGCTCC TCCCAAGCTT CAAATAGCT CTCAATGATG CTCACTTGT TGGAGAGAG 180  
CTCCGATAG GAGAGAGCT GCGCTGTACA GTCTCTTTGA CTGTATACGA GAGCTCTGTA 240  
50 TCAAGAGAC GATGAGAAA GTCCAGCTT AATGCCAGA ATTTGCACTT TCAACATGTC 300  
TGTTTGTG TGTGAGAC TGAATGCTT TATTTATTA CAGGAGTCT GATTTTTTTT 360  
TTTGAATGC TTTGTGCTA TATTTGTG GCTGTGAGA GAGAAATGAG ATTATTTGTA 420  
55 GATGAGATC CTTCATAC AGTACTTGG AAGCAGAC ACAGGTTGA AGACAGACA 480  
CGAGCTTGG AATATATAC TGTCAATGCT CTTTATATA AAGCTGTCC TCAACATGTA 540  
60 CAATATGCT ACATATGCTT AGTCTTTT TGAAGCTT TTACAGAG AATGTAGCT 600

WO 98/39448	PCT/US98/04493	461	462	PCT/US98/04493
5	660	720	780	840
10	900	960	1020	1080
15	1140	1200	1260	1320
20	1380	1440	1500	1560
25	1620	1680	1740	1800
30	1860	1920	1980	2040
35	2100	2160	2220	2280
40	2340	2400		
45				
50				
55				
60				
(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 458 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear				
(2) INFORMATION FOR SEQ ID NO: 235: (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 235: GGGTCGAGGA ATTCGGCAG AGAAGATGTT TCAATTTCTT TCAATTTTAA AGAATCTTCT				
30	60	120	180	240
35	300	360	420	480
40				
45				
50				
55				
60				
(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 591 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear				
(2) INFORMATION FOR SEQ ID NO: 236: (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 236: AGGATGAAGA GGAATATTAT TCTTGGATGT CTCTCCAGGA AATCCTTCTC TATACATTAA				
60				

WO 98/39448	PCT/US98/04493	461	462	PCT/US98/04493
5	660	720	780	840
10	900	960	1020	1080
15	1140	1200	1260	1320
20	1380	1440	1500	1560
25	1620	1680	1740	1800
30	1860	1920	1980	2040
35	2100	2160	2220	2280
40	2340	2400		
45				
50				
55				
60				
(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 458 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear				
(2) INFORMATION FOR SEQ ID NO: 235: (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 235: GGGTCGAGGA ATTCGGCAG AGAAGATGTT TCAATTTCTT TCAATTTTAA AGAATCTTCT				
30	60	120	180	240
35	300	360	420	480
40				
45				
50				
55				
60				
(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 591 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear				
(2) INFORMATION FOR SEQ ID NO: 236: (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 236: AGGATGAAGA GGAATATTAT TCTTGGATGT CTCTCCAGGA AATCCTTCTC TATACATTAA				
60				

TTAAGCTCT AGAATGCTA TGAAGGCTA ACTGACGCA AATTGACTT GATCTCTCTG 180  
AAATTAAGCT GCACTGACA ATGAGAGAG AAGAGAAAA TGTGATGCA TGTCTCTCC 240  
CAAGGATAC ATGAGCTTTG GATTGCTTT GAAATTTTT TCTTTTTTC TTCTCTCTCC 300  
TTTATGACAG TTGAGGACAT TGGAAATAC GAGCAACTC TCCACATCA ATGTACTCC 360  
ATGACATCT CTCTCTCTG TGGTGTATC TAAATTTGT GATAGGAAA CAATCTTT 420  
TGAATAAAA TAAATTAACA AACATAAAA GTTTATTTAG CCAAGCTGA GCTTGAAAAG 480  
TTTTTGTCA ATGACGAG AGAATACCT TTTTAAAGAG TACATATGT GAACTATAT 540  
GTACAGCTGA ATAAATTTGA AAGTGTGAT TTCCCAACT CTTGAGGAT T 591

## (2) INFORMATION FOR SEQ ID NO: 237:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1286 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

TTTATTTAG GTACAGCAG GAAAGCTAG AATCTGAG AAGAAACTG CCGTCTCATC 60  
TACAAAAGT GATTTACTT CTCTCTCTTC TTTTGTCAA AGTGGGCTTC CAGCAAGCAG 120  
GAAATTAAT GGGGCAATG ATGTATCTG TCAACTATA ACTATGACC GATTAAGAG 180  
CAGGAGAG GAAATGAGA AAGCAAGAT AAGAGCTCT TGTAAAGAT CTGTACTGA 240  
AGTAAGAAC AATTTACA AACCACTCC GTTTTCTCT CAGAGACTC CTCCAGCTCA 300  
CTTCAACT CTCTCAATTC TTCCAGCTC TCCAGCTGC AAGACTCTC CAGCTGTAT 360  
TTCCAGCACT GATTTCTCT CTCCAGCAG CCGTCAACT CAACTCTTA TACCAAGAT 420  
AGAAAGTGA CATTCTCTG GTTATGATG TGGTTTCCA CTTGATTTT CATTGCGAA 480  
TGGATTAAG AAGATTAAG ATACAGGAA TATGCAAAA GAGTTATGA GCTTCAGAA 540  
GCAAGTGA GAAATGAGA AAGCAATGA GAAAGCAAC AAGAGGAGA AGAGAAAA 600  
AGACATAGT CTTCGAGG TAAATGAGA CTTCCACTG AAGTGAAGA AGAATTAAT 660  
GACAGAAC ACATTAACA AAAATCTTAA AGAGGAGAG AAGAAAAA ACGGCACT 720  
GAGCTGCCC CTGAAAGGA GAGCACTGA CTTACACTG CAGATTAAG ATGCTTTG 780  
CGTTTGTAT ATATTAATC CAGAGTGA TACTTAAAT CTGTATTT TTCTGATTA 840  
TGTTAAGA ATTAACTTA AATCTCTTC TGTTTTATG TATGAAAGT TAACTTTTT 900  
TCCAAATTA AAGATGAT TTTTCAATTT AAGTTAAAA TCTTCTCTT GAACTATTC 960

AAAAATAAA AGCAAGCAT GACTTATAT CAGAGAAAG ATGTGAATG AGTCACTTAA 1020  
CAAGGATCT AAGAGCTCT GTTACTCTG TACATACA GATTATGCA GAAAGTCA 1080  
AAGGTTCAC TTGGGCAACA GTTTTGTGT TATCAACA CAATCTCTT AAGAGCTGC 1140  
ATCAAGAG GCAACAGAG GCGCCCTCT AAGCTTTGA GAAATTAAC TACCTTTAT 1200  
CAATACTGC TGTGACACT CTGACTTGG TATTTTGA GCGATCTCTT ACTTTTTT 1260  
GGTTTCAC ACTTTTGG TTGGGC 1286

## (2) INFORMATION FOR SEQ ID NO: 238:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 734 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

ATGCAAGCC AGAAGACA GAGAAAGT GCGAAGCCG AAGGCTGAG CCGACAGAC 60  
CTGTGCCAA ACTGATTC CTCCGATCA GCGGAGGAT GCTTGAAGC GCGCCGCGG 120  
AAGATCCGC CTGAGAAC CTTCGTGAC CAGAGAGCT TGTCAAGCC CCGAAGCTG 180  
GAGAGCTT GCAAGGCTT GTTACGAC GTGAGTACT ACGAGACA CTATGTCTC 240  
GTGTCTGAG GACTGATCT GTACTGTGT GTAGCTCC CTAATGCTT GGTGCTCTG 300  
GCTGCTTT TCGGCGCTT TTAATTTCT CTAATGCTT ACTTGAAT CCAAGCTGT 360  
CGCTTTGAG GAAAGTGA GCGAAGCCA TCAATGCTT TGGTGAAG GATCTCTTC 420  
CGCTTCTCT GACTGCTG TCGGAGCTG GCGTCTCTT GGTGCTGAG AACCACTG 480  
GTGTGATG GCTGCCAGC TCCCTTCA CAAATTAAG CTGTGAGCG GAGAGAGCTG 540  
CAATGAGAC CCGTGAAG TGTCTCTG GACTGCGG CTTCCGAGC CAGTGGCCC 600  
AACCTGCCC ATGCTGCTC TCCAGGCTT GCTGCTGAG CCGAAGCTC CTTCCATCA 660  
CAAGCCGAG GAGGATCC GCGTTTAAA ATAAAGCTT TATGAGCTC ATTCAAAAA 720  
AAAAAAAA AAAA 734

## (2) INFORMATION FOR SEQ ID NO: 239:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 809 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEINNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

5 CGGGGCTTTC AGGTTACCGG GTTGTATACA CGAGTCTTAC CCTCTACGAC GCACATACATG 60  
CAGGCGCAGAA GGAACACAGAG AAGATATCGG AGGCGGAGAG GTTACAGCGGC AGACCCCTGC 120  
TCCCGAAGCT GATTCCCTCC GTTCCAGGCC GCGAGTGGCT GGAAGCGGCC CGGCGCAGCA 180  
TCCGGCCCTG GAGCAGCTTC GTGACACAGC AGCGCTTCTC AGGCCCCCTG AACCTGGGAG 240  
AGCTGTGCCA GCGCTCTGTA CGCAGATGAG AGTACTACCA GAGCACTAT GTGTTCTGTT 300  
TCTTGGGCTT CATCTCTTAC TGTGTGTGTA CTTCCCTTAT GTTCTGCTAT GTTCTGGCTG 360  
TCTTTTTCGG GCGCTGTATC ATTCTCTATC TCGCGACCTT GAGTCTCAG CTTTGTCTCT 420  
TTGGCGCAGA GTTACGACCA CCGCATCATG ATGCTCTGCG TGGAGCATC TCTTTCCTCT 480  
TCTTCTGGCT GCGTGTGCGG GCGTGGCGCG TCTTCTGGGT GCTGGAGCC ACCCTGTTGG 540  
TCAATGGCTC CCACTGTGCC TTTCACACAGA TTGAGGCTGT GGAAGGGGAG GAGCTGCAGA 600  
TGGAGCCCTT GTGAGTGTTC TTCTGGGAGC TCGCGGCTTC CCGGGCAGC TGGCCGACCC 660  
CTGGCCATCC CTGTCTGACA CCGCTCTGCT GTTGGGGGCC ACAGCGGCTT CCGATCAGAA 720  
GCGCGCGGAG GCAATCCGCG TTGAAATAA AAGCTGTATT GCGTGTCAAT CAGGAAAAA 780  
AAAAAAAAA AAAAAAAAAA AAAAAAAAA 809

(2) INFORMATION FOR SEQ ID NO: 240:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2201 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEINNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

5 TCAGACACAG GTTCTGGGCA CATGCGCGCT GCGGTGTGTC AGGCGCGCGG CTGACGACCA 60  
GCAAGTGCAG GCGGTCTCTA CCGCGGTGTA GCGGTGGCTT CCGGTGGGTA TGTGCGCTC 120  
TTACGCGCGC TCTGTGTCCC GACATCAAGT GTATTTCGCA CTTCCCTCTC GCGCTGTGTG 180  
TCTACTGAGA CCGGAGGCGG TGAAGGCGCC CCGGTTCCTTT CTCACTGTGTC CTCTGTGCTT 240  
CAGGCGCAGC TCCCTCTCTC CCGGCGACT TCCCTGCGCT GTGTTGCGTC CATCTCTCTT 300  
TCTCCAGGCT CTTCCCTCTC CAGGCGGATG AMCGGAGCA CCGGCGAGTC CTTGGAGGCC 360  
CGGCTTTC AGGCTTCGAG GGGAAACCGA AGTCCGAGCA GCGCGAGTTC CCGAGCAGG 420

ARGGCTGCA GCGCATCAGC GCGCTGTCTC CCGCGGTTTC GCGTCTCTATA GTGGGCTGC 480  
TGTGCTACAT CATTCTCTCG AACTACATAG ACCGCTTAC CTTGGCTGCG GTCTTTCGCG 540  
5 ACATCGAGCA GTTCTTTCAC ATCGGGGACA GTAGCTCTCG GCTCATTCAG ACCGTGTTCA 600  
TCTTCAGTTA CATGTGTGTT GCACCTGTCT TTGCTACTCT GGTGACAGAG TACAAATCGGA 660  
AGTATCTCAT GTGCGGGGCG ATTGCTTCT GGTTCCTGCT GAACTGTGAG TCAATCTTCA 720  
TCCCGGAGCA GCAATTTCTG CTGCTCTCTC TGAACCGGCG CTTGTGTGCG GTGCGGAGG 780  
840 CCGATTTATC CACCAATTCG CCGACTCTCA TTGCGGACTT CTTTGTGTGCG GACGAGCGGA  
900 CCGGATGCTC AGCACTCTCT ACTTTCCTAT TCCGTTGGCG AGTGTGTGAG CTTACATTTGC  
960 AGGCTCCAAA GTGAGGATA TGGCTGAGGA CTGGGACTG GCTCTGAGCG TGAACCGCGG  
1020 TCTAGAGTGG GTGGCGCTTC TCTGCTCTGT CTTGTGTGAG GCGGAGCGCG CAGAGCGAGC  
1080 CCGTGGAGCG CACTCAGATT TCCGACCGCT GAAACCGAGC TCGTGTGAGG CAGATCTGAG  
1140 GCGTCTGCGA AGAAATCTTA GTTGTGTCTT GTTTCCTGCT GCGTTCAGTG CTGTGCGCTT  
1200 TGTACAGGCG TCCCTGTGCT TGTGTGCTCC GCGATTTCTG CTTGCTTTCG GCGTGTGCTT  
1260 TGGGAGAGCC CCACTCTGCG TTCCCGGAGA CTCTGTGCTT TCCCTGTGAGA GTCTCATCTT  
1320 TGGACTATC ACTGTGCTGA CCGGAGTCTT GCGTGTGCGG CTGGGTGTGAG AGATCAGCGG  
1380 CCGGCTGCGC CACTCAGAAC CCGCGGCTGA TCCCTGTGTC TGTGCGCAGT GCGTCTGCGG  
1440 CTCTGCGACC TTCTCTCTTC TGTGCTTTCG CTGGCGCGCT GGTAGCATG TGGCGACTTA  
1500 TATTTTATAT TTCATTGAG AGAACCTCTT GTCATATGAG TGGGCGCATG TGGCGGAGAT  
1560 TCTGCTGTAC GTGTGTATCC CTACCGCAGC CTGACAGCGC GAGGCGCTTC AGATGCTGCT  
1620 GTCCGACCTG CTGGTGTGAG CTGGGAGCGC CTACCTCATT GCGCTCATCT CTGACCGGCT  
1680 GCGCGCGAAC TGGCGCGCGT CTTCTTGTTC CAGATTTCG GCTCTGTGAGT TCTGCTCAT  
1740 GCTCTGCGCG TTGTGTGCGG CACTGCGCGG GCGACTTTTC TGGGCGACCG CATCTTCAAT  
1800 GAGGCGCAGC GCGCGCGGCG ACAGCTGAC GTGAGGCGCC TGTGCGAGA AGCAGGCTCC  
1860 ACAGAGCAGC GCAATTTGCTT GCGCGCAGCG GCGCGCTTCA CCGCGCTGCG CTTGCGCAT  
1920 GTGCTCATCT GAGAGCTGCG GCGTCACTTA CCGTCACTAT TCCGACAGCT GCGCTGCGCG  
1980 CCAACCCAGC AGGCGCTGCG GCGTCACTCC TTGGCGTGGC CAGCTTCCA GAGGAGCGCT  
2040 GCGCGCTGTC CAGCTTCCA GACATCTAT GGTGTGCTGA GCGGAGGAGG TGGGCGGTTCA  
2100 GCGAGGCGAT CCGTCTTCAC AGGCGCAGCG CCAAGGCGTC GGTCTATTTT GTACAGGAT  
2160 AAAAAATTTA GCGAGAGCGC AGGTGCTGCG TCTGTCTTTT CTCTGCTGCG CTTCTGATCT  
2201 TGAACCGCTT CTTACACCGA GCGCTCTGTA AGACTGTGCG T

## (2) INFORMATION FOR SEQ ID NO: 241:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1661 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

60 GTCTTCCCGG AACTGAGACA GTTCTGAC ATCCGAGACA GTAGCTTGG GCTGATCCAG  
120 AACTGTGCA TCCGAGTTGA CATGTGTG TGACCTGTG TTGGTACTT GGTGACAGG  
180 TACAAATGGA AGATATCAT GTGCGGAGG AGTACCTCT GTTCCCTGTG GACATGAGG  
240 TCAATCTCA TCCCGAGAGA GCAATTCTG CTGCTCTCC TGAACCGAGG CATTGTAGG  
300 GTGCGGAGG CAGATTATC CACATGCGG CCGCATCA TTGCGACTT GTTGTGACC  
360 GACAGCGGA SGGATGCTC AGCATCTCT ACTTGGCAT TCCGATGAGG AGTGTATG  
420 GCTTACATTC AAGCTTCCAA GTGAGAGATA TGGCTGAGA CTGGCATG GCTCTGAGG  
480 TTACACCGG TCTAGAGTGG GTGCGCTTC TGTCTGCTT CTTGTAGTGG CCGAGAGCG  
540 CAAGCGAGG CTTGAGGCG CACTGAAAT TCCGACCTT GAAACCGAG TGTGTATG  
600 CAGATTGAG GCTCTGCGA AAGAAATCTA GTTGTCTCT GTTTCCTGG GCTTTCATG  
660 CTGTGACCTT TGTGAGGAG TCCCTGAGTC TTGCGGCTCC GCAATCTGG CTGCGTCCC  
720 GAGTGTATCT TTGGGAGAG CAGACCTGCG TTCCCGAGAA CTCTGCTCT TCCCTGAGCA  
780 GTCTCATCTT TGGATCATC ACCTGCTGGA CCGAGATCT GAGTGTAGG CTTGGTGTG  
840 AAGTACAGG CCGATGCGG CACTGCAAC CCGCGAGCTGA TCCCTGTGTC TTGTGCACTG  
900 GCTCTCTGG CTGTGCAAC TTCTCTTCC TTGCTCTGG CTGCGACCT GGTACAGATG  
960 TTGGCATCTA TTATTTCATC TTCAATTGAG AAGACCTCTT GTGCAAGAC TTGGGCAATG  
1020 TTGGGCAATG TTGTGTATC GTGATGATCC CTACAGAGCC CTACATCATT GCGCGATCT  
1080 AAGATGATCT GTTCCAGCTG CTGGGTATG CTGGAGAGCC CTACATCATT GCGCGATCT  
1140 CTGACAGCTT GCGCGAGAG TTGGCCCTCT CATTCTTCT GAGATCCAG GATGTGAGT  
1200 TTGTGATCAT GATGTGAGG TTGTGTAGG CACTGCGAG CACTGCGAG TTGGGCACTG  
1260 CATCTTCAT GAGGCGAGC GCGCGAGGAG AAGATCTGAC GTGAGAGCC TTGCTGAGAA  
1320 AAGAGGATCC AAGAGAGCC GATTTGTGAT GCGCGAGAG GCGCGATCA CCGCGATCC  
1380 CATTGCAATC GTGTGATCT GAGAGGCTG CACTGCACTA CTGCAAGATC TTGCGACACT  
1440 KACCTGAGG CAGCTCAGG AAGAGCTGG GCTTACAGCC TTGCGCTGC CAGATCTCA

## (2) INFORMATION FOR SEQ ID NO: 242:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1146 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

60 TACACAGAAA GCAAGAGATG AGACTCTCT CATTCATCT TTCTTACGCC ATCTGTGCT  
120 GATCTTCCC CCGCCCATCA TACCTCTCC TTCTGTAGAC CTGTCCCGG TTGGCTGTAA  
180 TTGTGCACTT TACCTGATA TTTCATGAGG AGGATGAAA GCGAGATCA CCGTACCGG  
240 TTGGAGCATG GCGGAGAGGA AAAAGAGGAG GATGACAGG AAGAGGCTGC AGGACAAAC  
300 AAGAGGTTGG GGTGAGGAGA AAAATGTAGG GAGAAATGAG GTTGCAGGCC CTGCAAGCCG  
360 GTTTCAGCAG CAGCTGCGG CCGCCCGAG CTTGCGATC CAACTTCCGA GACAGGGTAC  
420 CAGCTCTGCT GTGTGTATCA TGAATTTCT TCAATAGTGG TTATGCACTA TCTTCCATAG  
480 GTTAAAGAG CCGTGTGCTG CACATGACA TTCCAGCCCT AATATATTAAG ATTAATATAT  
540 AATATATCT AATTAATTTA TGAAGATGAG GCGCTTCTG CCGTCTGGA CCAAGCAGAAA  
600 CTAGACCTTT TGGTGTGCT ATTAATTTCT GTTGTGTAT TTGTGTCTTT TTGTGCTTGG  
660 TTGTATGCTG TGAATGACA AGTGCAGTCC GATATCTCT GTATTAAGAA ATTAATTTTT  
720 TAAATTCATG AATGTGTATG TAAATGTAGC CTGAGACCT CATTCTAGCC TAAATTTAGG  
780 AAGTTGCCA ATTGTGTCT TTCAATTTA CTGATTAAT TTGTATCA AATCAATCT  
840 TTCTCTCTTT CTCTCTCTCC CACTCTGAC CATTGCTCT TCAATCTCC TTGCGCGCCC  
900 TCCCTCTCTC CATTGTGCT CCGGTGTAT TTCTGTAC TCAATCTCT CTGCGCTCT  
960 CTTGCTCTCT GATGCGCTCT CCGCAGCCA CTGCGCGAG TTGTGCTGC CAGCTCTTAT  
1020 CTGTCTTATG TTGGAAGAG TTTCATCTGA AATTTGAGG TTGTGTGCTC AAGTTTCCCG  
1080 ATCTGCTCTC GTTGTGTGTA GATTTGAGCG TTTCCTTTGA ATTTGATCTT TTCTGAGAG  
1140 AATTAATTA AAAAATAGAA AAAAATTA AAAAATTA TCAAGAGGAG GCGCGTAC

CAATTG

1146

5 (2) INFORMATION FOR SEQ ID NO: 243:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1350 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

15 AACCCAGGCG TCCTGGGCGA GGGCTGGAG GGCAGAGGCG CGCGAGGCG CAGTTCCAAA 60  
CATTGGCTAG AGCAGAGAGG GCGGAAACCC GTTGGCCGAG CCCAGCGAGC TTGACAAACC 120  
CTTTGACGCA CCACAGAGCT ATGAGCCTCC AGCCCCCTGCC CCATTTGCCTC CACCCTCAGC 180  
TCCCTCTGTT CAGCCCTCGA GAAGCTCTAG CCCACAGAA CTTAAGAACT ATGGCTGATA 240  
CAGCACTAGT GCGCTAGCTG CAGCAGCCAC AGCTGAGCTG CTGAGAAAC AGCAGAGGCT 300  
CAAGCCGAAG GCAGAGGAGT TGGACCGAAG GAGTCGAGAG CTGAGAGATG CTGCCCCTGG 360  
RGCAGAGCT ACTGACAGCA ACATTTGGCC CCTCTACTCT TCTTTTCTC CAGTTCCAGC 420  
CTGCTTTTTC CAGGACATCT CCATGAGAT CCCCCAGAA TTTCAGAGA CTGTATCCAC 480  
CATGTACTAC CTCTGAGAT GCAGCAGGCT GGCCTCTCTC CTGAGCTTCC TGGCCTGCTT 540  
GGCCAGCTTC TTGTGTGAAA CCACATGG GCGAGGCTTT GGGCTTTCTA TCTCTGGGTT 600  
CCTGCTTTTC ACTGCTGCTT CTTTGTCTG CTGTGACCG CCATGTATA AGGCTTTCCG 660  
GAGTGAGAT TCATTTCAAT TCTTCTTTT CTCTTTCAT TTCTCTGTC AGGATGTGCT 720  
CTTTGTCTTC CAGGCCAATTG GTATCCAGG TTGGGGATTC AGTGGCTGGA TCTCTGCTCT 780  
GGTGTGGCG AAGGCACAC AGCAGTATCC GTGCTCATGC TACTGGTGGC CCTGCTCTTC 840  
ACTGCAATTG CTGTCTAGG AATTGTATG CTGAAACGGA TCACCTGCTT ATACCGGCGC 900  
ACAGGTGCCA GCTTTTCAGA GGGCCAGCAA GAATTTGCTG CTGTGTCTTT CTCACAGCTT 960  
GGGTGTGAAA CCGCAGCTG CCATGCGAG CCGTGGGCT CTGCAAAATG CCTTCCGGGC 1020  
CCGCTGAGCC CTGACTGGGA TGGCTGGCC CTGCTACTTG AGGGAGCTGA CTTAGCTCCC 1080  
GTCTCTAAGG TCTCTGGGAC TTGGAGAGAC ATCACTAACT GATGGCTGCT CCGTAGTCTT 1140  
CCCAATCTTA TGGCCATGAC TCTGTAACT GACAGGGTGG TGGGAGTTTC ACTGTAGCTT 1200  
AGTCCCCGCA TCAGGCCACA CTGCTCCAC CTCTCAGAG CCCCAGCCA GCTTCCCTCT 1260  
GCTGTGGCAC GGTGTGCTT TGGTTATTT AAATPAAAG AAGTGGAC TGGAAAAA 1320  
AAAAAAAA AAAAAAAG GGGGNCNC 1350

5 (2) INFORMATION FOR SEQ ID NO: 244:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1529 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

15 TCCAGAGGC CGGGGGTTC CAGCTCTGCC TGTAGCAGAG CCTGAGGAG GAGGAGGAG 60  
AGGATGTGCT GAATATACCTC CGGAGATCT TTTTCAGCTA GGGATTAAC TGTGCATCTA 120  
ACTGTCTGCC GAGAGGAGCT GAGAGGAGC TGAGCTTTCA CTGTGTGTC TGGGCCGACC 180  
GCTGTGTGGA ATGGGGCTCT CTGTGCTCTT ACCTTTGTGC CTTCTGTGGC CTGCGAGATT 240  
CAGCTCAGGC CAGAGGCCCC TGAGACTTCC GGGCCTTGGG GTGCCGTTCCT GAGTGTGGGG 300  
AAGGCGAGAC TCMAAATGAG ATCCATTTTG ACTGCTCTG TATGTACTGT GGCCTCTCTT 360  
GGCTCTTGAG GCTGTGGAGT CCAATTTGTC TGTGTTAGTC AGTACCCAGG TTCACAGGAA 420  
AATATATGCA TGTGTGCTC CACTTACTG GAACCAAGA GACAGTACTT TGCAAAGAAA 480  
AGGATCACTG CGAGTGCAC TGGATTTGCT ACAGTTTACT CCGCATGTC TCTCTGTGAG 540  
GAGGAAGGCT GTTTCAAAAA TAGTTTCCAT CATTGATCTA TCATGAGCT CCCACCTCTC 600  
CAGCAGGCTT AGAAGCAA CAGCTGGCC ACAGTTCTCT GCGCTGTCTG GAGGTGTGAG 660  
GCCACAGTGT ATAGACTGGT AAGCCAGACA GGCCTCTGCC CCGAAGTGC TACCTTGTCT 720  
TCACCTGTAC CTGTGTCCC GGGGAGCTAG CTTTAAGCA AGAGGAGAG GAGCCAGAAA 780  
GAGACACTGA GGACAGAGA TCACACCGA GTACATGTCT CTGCTCTCTT TTTTCAGTGTG 840  
GCTTTGGACA GGAATATATG AATAAATCAC TCCATATACG GTTTTCCAAI ACAAAAGTCC 900  
TAGAAAATAC AACAATTTCC CCAATGGCTA AGTTGTGCTA ATGTCTTTTC AAGTTCTGGG 960  
TTGGGAAGTG GAGGGTGGCA GGGTTTGTCT GTGGCAGCC GTCCAGTCTCT GTTCACAGCG 1020  
AGGATTTGGA GTCTCTCAGG GTCTCATCAT GGGAGTCAAT TGTACAGCGA CGCCTCTGCC 1080  
CTGTCTGCTT TCAGGTTCAG GGAAGCTTTG AAGCAGTCAA GCGTTGTCTT TGTATCCCAT 1140  
GTGTCTCTTC TTTTGTGAGT CACTCAGAGA TCACCTCTGG ACCCTCTGGG TTGGAGTTCC 1200  
AGTGAATGCT TATGGCGGCC CACTCACTAT GGTGGGCTGA GTGGAAGTC CTTAAGCAATG 1260  
TCCCAGAGA CACTGAGTGT CTGCTCTTTT TAAATGCTTC GTTTGTGGC GTAAAGTTCTT 1320  
TGTAGGTTT CATTTTGGCA TTGGCAAT CAGCCTGAAA GTCTGCCCC ATGACAGCAA 1380



1440  
1500  
1529

10 (2) INFORMATION FOR SEQ ID NO: 245:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1537 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

20 GTCGAGAGTC CCGCGACCC CCGACGCGCC TTCCCGCCGC GGGCGCGTCC CAGACGAAAC  
GAGCGCCCTCG AGACCTCCAC CTACCTTCACA GCGATGAATC CCGACGACGA ACTCGGAGTC  
180 AGCATTCGCT AAGCGCCGAG TCGTTGAGC TCCCTGTATTA ATCTGTAGC TGTCTGTGAAA  
240 TCGCCCTGAAA TTTCACCCCT CAGGTATATC TTCCAGTTAC ACAGATTCCT ATCGAGATGCG  
300 TTGTGAGCAT TATCTTACTC TATCGAGATA TGTTCAGCAT TTTCGAAATC ATCTTACAGA  
360 GCGAGCTCGC AGTTTGTAAA CTGAAATTTGA AGAGTTTGA CAGACCTTGA ATGCTTGTCT  
420 TACACAGCAT GATCTTTTTC AAGAACTGT GAAACTGATC TATCGAGAGG CCGAGATCAT  
480 CCGAAATTTT TCTTATATCG GAGCTCCCTT CTGTATATTC CTGTCCGATC ATCTGACAT  
540 TACCCGACAG AGTGGGAGCT TCGCGAATT GCTACTTCAA AGATGTGGA CTGAATATGA  
600 AGTAAAGAT CAGCTGTCAA AAGCGATGA AGTATCTGA AAACGATTC ATCGATTTGT  
660 ACTCTTCTCG GAGAACTTT ATCTTAACCT GAGATGAGAG GAAACAAAT CAGCGGTTAC  
720 AAGACAGAT ATCTTTCAGG TTGTCTTTCG AGAATTCCTG AATCCCTCTT TTCTTAAATC  
780 TATGATGAC AATTTAATTT GTTCAGTTAA ATTGTAAAG TTGACAGAT CAGTTTGTGA  
840 AGATCTTCG AAGGAAAAAG GAAAGATGA TATCGAGAGA ATTATTCGA GAATTCGAAA  
900 CCGTCTCTTA GATCGAACT GCGATGAGGA TGTAAACAG ATGCTCTGA AGCTTTTGA  
960 ACTCGGCTCA AGTAACTCG GCGAGATCA TCGAACTCA ACATATGAG AAGCAGACG  
1020 AAAAAATGAT CTTAAGTACT TTATGATGA ACAGAACTT TATACATCG ATGCTGTCC  
1080 TTTCATCTGA GCTGATCGAG ATTACAGGA GAATTCACA GAATTAAGTG AAGAGAGGA  
1140 CTTTTCCTCA GATTATGAG AAATGAGAC AGATTATCC GGGCTGTG ATCCATACT  
1200 GATGATATTT GATGATGAG TCGACCGGA GATGAGAGGA CATTGAGAAA AGTTTGTCTT  
1260 GAAATCGAGG CTATACGGA AACGATGAG TTAAATTTCA GCAATCACT TTATTAAGC

1120  
1180  
1440  
1500  
1537

15 (2) INFORMATION FOR SEQ ID NO: 246:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 506 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

25 TCGAGATTT GCGCGACCC CCGCGCGTGG GCGGTCTTA TCGCTTCCA GAACTGATC  
60 AAGCAGCGAG CTGAGAGAG TTGAGGAGAA GTCTGCTTC TGGCTCTGCA GAGCGGATCG  
120 ATTAATCTGA CCGGAAATTA AAGCATGCC CTCTCTGCTT CAGTGTGAAA GCGCAGCTGA  
180 AGATCTCGG CCGGAGATTT ATCGACTGAC TGTATACAC AGTATTCATG CTGATCGTAT  
240 CTGTCTTCG ACTGATACGA GAAACGACA CATGAGAT TGGTGAAGG GTGTGTGAC  
300 TTGTGACAG AGTATGCTGT CTTCGACAG GGGCGCTTAT TTACGAGAG CTCTCTTCA  
360 ATCCGACCG TCTTACGAG AAAAGCTCG TCGATGAAA AAAAGATTT TTGTATTTT  
420 ATTTACTTT TTATTTGAT ACTAGATAT AAACATATTT CTGATTAAT CCAAAAAAA  
480 AAAAAAAA AAAAAAAT TTGTGCG  
506

45 (2) INFORMATION FOR SEQ ID NO: 247:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1348 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

55 GTCTTTCCTT TACTGTTTGG AGTGTGAG TGAATGATA GGTATCATG GCGCTTCAG  
60 ATGACCCCTT GAAACTGCG GAAATCTTT AAATTCGAG TGGATCTCG GACCTGAGAG  
120 GCGCTTGA GCGCAGCTC TCGAAAAAC TGGAGTTGA TCGCGAGAG TCGGAGAGAG  
180

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

TCTGCTGAG GGCAGGCTGC CTTGACAC TGGTAGTTCT GGGCTGCGA GGGAGGGG 240  
CTCGGCTTT CTCTGAATG AACACTGCTC TTGACGAGTT CAGTACTTG TTCTCAAAAC 300  
ATTTCATAT TGAATGGTAG GTTTTCATTA GCATTTGTTT TTTAAGGCAT GGAAGGGA 360  
GATGCTCAA GCAAGTCATG TTTGTTTCA GTGGGATGG CCGGTTTCT CACTGCTGG 420  
GGCTTCCCT TGCATGTGCG ACCTTTGTGC AGGGCCACCA GGCAGACTCT TCCACTTC 480  
TCCACTGAA GCACAAAGG GCTTGAACG TAAITGGCT AATCGAGGC ATTTTITTTG 540  
TCTTAGTATC TTTCACATTT GTCCACGCT CTATTTTITTT TAAAGTTTCT GTTCTCTGTA 600  
TTAAACAGAA ACTAGAGAGA AATAGTTTCT GAAGCCAGTT TATGTCGAG ATCCCAAGG 660  
GAGGTTCCG TAGAGAAAA TAGTAACTG GTTTAGAAAC TGAAGAGGC AAACAGCCAG 720  
GAGCCATGG AGAGCAATTT GCAAAAGATC TACCTGACA TAAAGCCTGT CAGTGTCTTT 780  
CACCAGTGA ATAAACAGCG CTCCAAAGTG TTTTCTCTCT TCAAAAAA AAATTCACA 840  
AGCTTTTAAA GGTGCATTTA AGATCCATG TCACTTTTGA ATGAACTGC GGGCCCTGCG 900  
AACTGTACG TGTGTAGAA GCTTCAAGC CTCTGCAATG CATGTGATAC TCACTCTCAT 960  
TTTGTTCCT TGAITGCAT TTTGTTCTTT TAGCAGATCT GTCCGTGTG GTGCTGCTA 1020  
AGAAATGGA CACTTTGTT TTTGTGTTAG ATTGAAGCTG GCACTGCAA TCACTTTCTT 1080  
TATATGCAA TTAGGACGA CCAATCTGT GTTCCCTGCT TGTGCTTAA TGAAGTCAG 1140  
GGAGGAGG ATGTACACCC AAATGAGCG CCTCCATTT GCTTTGCCA GGCAGACAC 1200  
TTCACATCT TTACATGTT CTGTGTAAT TTAAAGTTTA TGTGTATPAA GCAAGGCTGT 1260  
TTCTGTCAA CTGTATATT TGTAAATPAA TATATGCTA CTTTGAGAT AAAAAAAA 1320  
AAAAACTGA GGGGGGCTCG GTACCCA 1348

(2) INFORMATION FOR SEQ ID NO: 248:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1766 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

55 GTCCGATC GGCAGAGCG CACAGGGCG CACAGAGCA GCGGACTTAA AGGACTTGA 60  
GCGAGCAAT TCCCGATTA TTCTATTCC CTTCCCTCTC TCCGCCCCG TATCTCTTTT 120  
CACCTTCTC CCAAGCTCG TCGGTASCA TGGCGAGCG TCGCGGCCA CTCAGTCCA 180

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

TTCCATCTG TGTGTGCTT TCGGAGCCGA GCTGTCCGG CCGGCGGGG GCGGAGGCG 240  
AGAGCTGCG CCGGCGCTGG GGCAGAGAG CTCAGCTCC TCTGTGCGG TGCAGATCT 300  
GATTTCTCG AGAGATGGA AGAGACTGG GTTTGTCTTT GGCACACGC TGAATCTCT 360  
GCTTTCCTG GCACTTTCA GTGTATCAG TGTGTTTCT TACTCATCC TGGCTTTCT 420  
CTCTGCACC ATCAGCTTCA GATCTACAA GTCCGTATC CAGCTGTAC AGATCTGGA 480  
AGAGGCGAT CATTCAAG CTAAGCTGA GTTAGACAT ACTGTGCTT CAGAGCTTT 540  
CCTAATTAC ATCAATGCTG CAGTGTGCA CATCAACAG GCGCTGAAC TCAATTATCG 600  
TCTCTTCTG GTAGAAGATC TGTGTGACT GTTGAAGCTG GCTGTCTTCA TGTGCTGAT 660  
GACTATGTT GGTGCTGTT TTACGGGAT CACCTTCTA ATTCTTCTG AACTCTCAT 720  
TTTCACTGT CCAATGTCT ATGAGACTA CAGAGCCAG ATTGATCACT ATGTTGCGAT 780  
GCGCGAGAT CAGAGCAAGT CAATTTTGA AAGATCCAA GCAAACTCC CTGAAATGC 840  
CAAAAAAG CAGAAATAG TACATGAAA CCGAATGCG AACAGTTCT AAACACCAT 900  
TTAATAGTA TAACTGCTT ACTGTACTA TGAAGGAAA TACTCAGTGT CAGCTTGAAG 960  
CTGCATCCA AGCTTTTITTT TTAATTTGT GTTTTCTCC ATCTTTTCC TTTAAGCTC 1020  
AGTATCAGC AAAAAATTT ATGAGCTAT AAAGACTTA TCTTAGACT CAGAGAGGA 1080  
AAGATCAA TTCAATGAT AAGTCAATAC CTTAATGCTG TGAAGCTT TACTGTAGC 1140  
TTGAAAGCG AAGATTTGA GTTAAGAGG AATATGAAG AACACTCTG GGTCTTCTG 1200  
TCCAGTTTC AGCACTAGTC TTACTCAGCT ATCCATTATA GTTTGCGCT TAAAGATCA 1260  
TGATTAATT ATGAAAAAT TATTTGGGA CAGGATGAG ATACTTCTT TGGTTTTTT 1320  
TTGAGCCTT CAAATCTAT CTCTCTGCC CACAATGGA GCAAGTACC CTCATACTCC 1380  
TTTTCTTTA TGAATTAACT ATCACTTGA TAAATAACTT ATAGTGTATA GTCAATATC 1440  
CTGATTCGA GATGCCATC TGAATAAAA GAATGAGAT GGAAGTGGG ACTGAGAGG 1500  
AGTCAGCAG CACTCTGCG TGGCGTCA CTTCTGTCC ACTATGCCA GCGAAGGAA 1560  
RGCTGCGCA TTTGGGAAAG TGTTTTCTAC GTCACTGAC ACCGTTCTG AGCATTAAT 1620  
TGAAACTCG TTCCGGAATG TGTTTTCTC CTTCTCCCT GCGCACCTCA AGTTAATTA 1680  
ATAGGTTGT ACTTTTCTTA CTAATAATA AAAAAAAA AACTGAGGG GCGCGGCTA 1740  
CCCAATGCG CGATATCAT GTAAA 1766

(2) INFORMATION FOR SEQ ID NO: 249:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2664 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

AGTGTCTCCG GAGGAGGCGG AGTAAAGGGA CTTCAGGAGG CAGATTCCG GATTATCTTA 60  
TTTCCCTCCG CTCTCTCCCG CCCGATATCT CTTCACACC TTCCTCCAC CTCGCTCCCG 120  
TACATGAGCG GAGCGTCCGC GGCACCTCAG TCCCATTTCA TCTCTCTGTC GTCTTCCGGA 180  
GCCGACCCCT CCCGCCCCCG CCCGCGCCCG AGCCGAGGAG CTCGCCCCCG CTCGCGGAGC 240  
AAGAGCTGCA GCTCTCTCCG TCCGCTCCAG GATCTGATTT TCTGAGAGGA TGTGAGAGAG 300  
ACTGGGATTTG TCTTTGGGAC GAGCTGATTC ATGCTGCTTT CCCCTGGGAGC TTTTCACTGTC 360  
ATCCAGTGTGG TTTCTTACCT CAGCTCGGCT CTTCCTCTCG TCAAGCTCAG CTTCAGGATTC 420  
TACAGTCTCG TCAATCCAGC TGTACAGAGG TCAAGAGAGG GCGATTCATT GAAAGCTTAC 480  
CTTCAGCTAG AGATTACTCT GTCTTCAGGA GCTTTTCATTA ATTACATGAA TGTCTCCAGT 540  
GTCCACATCA AGAGGAGCCCT GAAACTCATT ATTCGCTCTC TTCCTGGTAG AGATCTGCTT 600  
GACTCTTGA AGCTGCGCTC CTTCATCTCG CTGATGACTT AGTGTGTCC TGTTTTTHAC 660  
GGAATCCACC TTCTATATCT TCCCTGACTC CTGATTTTCA GTGTCCCAT TGTCTATGAG 720  
AAGTACAGAA CCGAGATTGA TCACTATGTT GGCATCCGCC GAGATCCAG CAGATCCATT 780  
GTTCGAAAGA TCCAGGAGAA ACTCCCTGGA ATGCCAGAAA AAAAGGAGGA AATAGTACAT 840  
GGAATCCAGA AATGACAGAG TTACTATTAAC AGCTTTTAAT AGTTATAGAG TGTCTACTTG 900  
TACTATGAGG GAAATATCTC AGTGTGAGCT TGAAGCTGCA TTCCAGCTTT TTTTTHTAAT 960  
TTGTGTGTTT CTCCCATCTT TTCCCTTTTA CCGTCAATAT CAGCCAGAAA AATTGATGGA 1020  
CTGATTAAGAG AACATCTTGA GAACTCAGAA GAGGAGAGAA TCAATTTGAT AGGATTAAGTC 1080  
AATATCTTAA TGTGTGTAGA GCGTTTACTT GTAGCTTTGA AGGGAGAGAA TTGGAGGTAA 1140  
GAGAGAGAAAT GAAAGAGAAC CTCGTGGTCC TTCTGTCCAG TTTCCTACAG TACTGTACTT 1200  
CAGATATCCA TTATATGTTT GCGCTTTAGA AGTCAATATT AACTATGAA AAAATTAATT 1260  
GGGAGAGAGA GTGTGATAGC TTCTGTGTTT TTTTTHTGA GCGCTCAAT CCAATCTTCC 1320  
TCCCCAGAA TGTAGACAGC TTAAGCTGAT ACTGCTTTTC TTTHATGAT TTACTTCTGAA 1380  
CTTGATTAAT AACTTATAGG TGAATGATAT AATTCCTGAT TCCAGAGATG CCAATGATTA 1440  
AAAAAGATA GAAATGAGAA GTGGGACTGA GAGGAGTCA GAGAGATCC TCCGCTGCGG 1500  
GTCACTCCCT CTCGACTAT CCCGAGGAGG GAAAGATCT CCGCATTTGG GAAATGCTTT 1560  
TCTACCTCAC TGGAGACCGG TTCTGACAT TATTTTGAAG ACTGCTTCC GATGTGCTTT 1620

TTCTCCCTCT CCCCTGCCA CTCGAAATTT AATTAATAGG GTTGTACTTT TCTTACTATA 1680  
AAATTAATGT CTGTACTCAG TGTGACTTCC TGTAAATTTG TTGAGAGAA AATTAATCTG 1740  
CAGTGGGACT CTCAGATTAAT TGAATTTTGG TGAATCTATC TGAATCTGCG GCGGAGCATT 1800  
CTCAGAGGCT GAAATACAGA AAGCTTTTTC TTCTGATCT TTTCGAGAA TTCAATCTTC 1860  
GATTTCCAT TTGGGGGAGG GTTTTTCCT TACATTTGAA TATGAGATTT CAGCGAACTT 1920  
GAAAGAGAAA TCAATCTGCA GTTCTCTCAG GTTCTGACTC AATGCTATCA TCTCTTCAGG 1980  
GGAATATCCA CTCGAGATTT TAAATTAAG GCTATGATAT TGAATGCTCC GAAAGTACGG 2040  
CAGCTCCAAA AAGTATGCGA AGGAAATGCT CTAATGTCTT TCGAAAATTT AGTTAGAGAT 2100  
TTGATAGGCT AAAAGGTACC CTTCGCTTAC TCCATCTTAAT TTTCATACCC CCGTTTGAT 2160  
GTTTTACTCG GTTTCATCTC CTAGTAGGAA GTGCATCTC CAGTCTCATC CTCTGACCTC 2220  
CCAGGAGATC AATGATGTTC TTTTGGGCTT TCCCTTCGAA AGGAGCTCTC GCGATGAGAG 2280  
TCCACATCCC AGTCTCTTTC TTTTGTGCTT GCTGTGTTTA GATTAATGAA GAGATCTTTC 2340  
TCCACACAG GATTTTTCCT TTTTTHAGA AAAACCTATA GATGAAATTT TACTATGAAA 2400  
ACTGTGTGTA GGTGTGTCTG GGTGCACAT AAAAATAGAG TGAAGCTTGA GAGATCTTGA 2460  
TCTTGTCTCC TGTAAATTTT CAAATGATG TGTATTAAT AAAAATGAAA AAACAGAAA 2520  
AAAAAAGAAA AAAAGGCGG CCGCTTCAGA GATTCACAGC TTACGTAACC GTTCATGCGA 2580  
GCTTCATAGC TCTTTCTATA GGGGTCCCCC AATTTCAAT CAGGCGGCGG TCGGTTTAA 2640  
AAGATCTCG AATGAGGAGG ATCC 2664

(2) INFORMATION FOR SEQ ID NO: 250:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 865 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

CTGAGGATGG AATGACAGG TTCAAGCCAT TTGACCCGCA CCGCTCTATCT CTCGAAATCC 60  
GGTGTCCCG GATTAGAGTC CCGATTCCTA AGGTGGGAT GGTGTCTCTC GGGATGAGAT 120  
TTGGGCTTTC CTCGAGCCCTT TGTGTGATTC GGTGTCTCTA GATGAGATTT TTAGGTTTCC 180  
TCCGGGCTTT CCGATCTTTC AACTTAATATC CCGACTCCAA GATGAGAGAA GCGCGGAAAC 240  
TGAAGAGCTT GATTAAATAG GTGCATCAC TGTGCTTCTC AGGTGCTTGG GCGATGCGAA 300

TTGGGTGAC CTGTGCTCA GCTTCTCTG TTTTGGGAG CTTTCCCGA CATACCTTGG 360  
GACTAGTGA GAGCAACTC TTCCCTCTCT ACTTCCACAT CTCCTATGGC TTGCTCTTCA 420  
5 TCACCTCTG CATCTGGCT TCAGAGATG CTGTGGCTCA GCTCACAATC TTGGAGGCCA 480  
GCCAGCTTTA CTTGCTGTTC CTGAGCTTAA CCGTGGGCAC TCCTAGAGCC CCGTGGCTGG 540  
AACCCGCAC CACAGCTGCC ATGTGGGCC TCCAAAAGCT GAGAGAGGAG CAGAGCCCTGG 600  
10 GTGGGAGGT ACCAGGAGC CACAGAGGTC CCGATCCCTA CCGCAGCTG CGAGAGAGG 660  
ACCCAGATA CAGTCTCTC CCGAGAAAT TCCTTCCCTA CCAATGGCTG TCCTCTCTTT 720  
15 GCATCTGGG CTGCTCTCG AGCAATGGG CTCCTCTCG TCCTCTCTG CCGAGATTA 780  
GAGCCTCTA GCATGGGCC TCATGCTTAA TAAATCTTC TTCAGAAA AAAAAAAAA 840  
20 AACCTGAG GGGGCCGCT ACCCA 865

(2) INFORMATION FOR SEQ ID NO: 251:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2082 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

TCGGGGGCGN ATGGGTGTC TGCTCANG ATGGCAAT CTGGAAATC TCATAACTT 60  
35 GCTAGCTGT TTTTCTTTT TTTTCTTACA CCCCCCGC CCAGCCCGG ACTTGCAAA 120  
TTCTCAATCA TCTCAGAGA GTTCTCATG TGAAGGTG ATCAGCTTG AAGCTGCAT 180  
40 CATTACATA TTTTCTCTC TCTTCCCT TCAGTTCATG AACTGTGT CATTTCTGT 240  
GTGTGTGT GTTTTATTT GTTGTATTT TTTTCTTAA TTTTCTTT AGAGCTTCT 300  
GTGTGCCA CTTTCTTCC AACTCCACC CTCAGTCTT CTCAGCCAT CTCCTCCAG 360  
45 ATGAAGAAA AAAAAAGCA AGTCTTTT TCTCTCTCT GAGTCTTCA TGTGAGATG 420  
AGCTTCAAA GGAAGAAA ATGTCAATG TTTATAGCT GCAGCTGCC GAGTTCATC 480  
50 GCGTCTTTT TTTAGCATG TTAATCTAA ATAGAGAAA AATGTCTAT GAATTCOA 540  
CAATCAAGCC TCATCAACC TTTCTGGTGT GACTTGTGAG TTTTGGCTT GTGATCCA 600  
660 ATCTGAGGT TTATCTGCC ATTAAGAAA CTCATCTCA TCTCATCAT TATTATGCTT 660  
55 GCTACTTGT CTTAGCAACA ATGACTATA ACTGTTCAA AGACTTATG GAAAGGAC 720  
ATTATATTA TAAAAAAA AAGCTGCAT CCTGACATG TATGTATTA TTTTCTTT 780  
60 CTTTCTTTT CTTTGTGCT TGAATGCA GCTTCGAGA CTATATGAT GCATTCATA 840

CTTTCTTTT ATTGCTCAT GACTTTTGT AGTTTAAAC AAACAGTCC AACCTGAG 900  
CTTCTGCC ATGAATTTT GCATCTCTC CAACCTGCT TTGAGTACT CAGACTTCA 960  
5 AACTCCAT GCAGTAGG CATCTCTGT GCAAGTATC CAGAATGGT TACACATTA 1020  
ACTGCGAA CATGAGAA CTCCTATGT TTTCTTTTA ATGAAGTGA CCCCCACTT 1080  
10 TGGGACTTA AATCTCTA TTCCGAGAA GCAGCTTAA ATTTATTTT TAAAGAGCA 1140  
AAGTCCCA TTTTCTGT TTTTCTTAT TTTTCTTTA TTTTCTTGG CTTTCTGCTA 1200  
15 TTGCAATG TGAATCTC TGCGTTCTA GTATTAAT TAAATCTAGT TTTTATATC 1260  
TGTAGCCCA GTTAATATGT ATCTCAGA TAAAGATG TTTATAGTAA ATTTGAAAG 1320  
GTTAGTCTG TTTAGCTGA GATTTTATA ACCATGATG CACTAAATG TTTACTATG 1380  
20 TCAATTAAG GGGGTAGAG TTTGCAAGG GACTGTATA AAAAAAGTAC TTATACAGA 1440  
TGTCTTGA ACTTAATAT AGTTGGTA TGTGATCT TGTCTATCC ACTGACTGA 1500  
TTGAAACCA AAGTATTAAG AGCGAAGC CCGCTGTTA TATCTTAGG GGTATTTAC 1560  
25 ATCAAAAT GTATGTTT TTTCTTTC AAAAAAG TATTTGGAC TGAATTCAC 1620  
TAAATATA CTTGAGCA TATATACA AAAAAATG CAACCTGTT TAGAACCTA 1680  
30 AAAAAATTA TCAATATA AATGTCAT TACTCAGC TTTTAAGATG ATGCAGATT 1740  
TTTACAGTT GTATGTGT GCAGACTG ATTTCTGTA ACTTAAAAA AATTCACAG 1800  
1860 TTTTAAGGC AATATCAT ATATGTTT TTGAGGACT GACTCTCTT CTTTAAAAAG 1860  
35 AATGAAGG TAAATCTAC CACAATAT ATAAAAAT CTTGTCAGTT ACTTTCTTT 1920  
TACATATTT CCTGTGAAA ATTGTATAT ATCTGAGTT ACTAATAAC CAGCGCTGT 1980  
40 GTTCTATGT CTTTCTTCTT CATTTCAAT TCTGTGATA TCAAGAAAG AATATCTAC 2040  
2082 AATATTAAC GCATTTTCT TTTCAAAAA AAAAAAAAA AA

## (2) INFORMATION FOR SEQ ID NO: 252:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1482 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

CAGCAGGCT GCGCCCGGG ACTTCTCTT GCGCTGCTC CCTCGAGGC CTCGCCGCTT 60  
GCCGCCCTGG CCGTAGCGA GTCTTAGGC AGATGAGG CTGTGTGAA CTTGTACCA 120

GAAGTGAATGA AACACGACAGA TCCCCGATAC CAGGGCTAAC CTCTGAATGGA GTCCCCCTTG 180  
CTAAATGAACT GAAATTCCTT GAGCTAGATG TAACTTCATG TCTGACTTGG GCGTCGACATC 240  
ATGGCTAAATC GAAAGCCCTT CAGATCCCTT GACTTCATGA TTGTCTACAA CTTCCTACATG 300  
GTGGACATCT CCGCTGACAT TTCTGAATGA TTCTGAATCT GCGGCTGGCT GAGACATCAT 360  
ACCTGGGGCT GTAGACCTCT GAGATTAATCC AACAGCCCTG AACGACTTGA GATGGTGGAG 420  
GTGGCTTGGC TCTTCCTCTT CTCCAAATTC ATTGAAGTGA TGGACACAT GATCTTAAAT 480  
CTCCGAAAGA AAGACGGGGA GGTGACCTTC CTACATCTCT TCCATGATCT TTGGCTTCCC 540  
TGGAGCTGAT GGTGGGGGCT AAGGATTCGC CCGGAGAGAA TGGGCTCTTT CATTGCGCATG 600  
ATTAACTCTT CAGTCGATCT CATTAATGTAC CTGTACTAGG GATTATCTGC CTTTGGGCGCT 660  
GTGGACAAAC CTTACCTTTG GTGGAAAAAG CAGATGACAG CATTTCAGCT GATTCGAGTTT 720  
GTCCGTGCTT GACTGACAT CTCCGAGTAC TAACTTAAAT CAGATGTGAA CTACACAGTAC 780  
CGATCATTTA TTGACCTCAT CTGAAATGAT GCGACATCTT TCTTCATGCT GTTCTCCAAC 840  
TTCTGGTATC AACTTTTATC CAGGGGAGAG GAGCTGCCCC GTGCACTTCA GCAAAATGGA 900  
GCTCGAGTTA TTGCGAAGAT CAGGCGCAC TGAAGACAT GCGCTGATTA GCGCGCCGAC 960  
TAAAGCTCTC AAGATGACAC CTTAAGGACAG TTTCCTGACAG TCCCTCTTCC ACCTACAACT 1020  
GTGACAAAG CTATGTGTGT CAGGACTGAG CAGGGGACTG GCGTCTCCCT CCGCAGACAT 1080  
GCTCTACAG GACCAAGCTT TTGGTCTCTC ACCCACTTCC CCGGGGACAC TCCAGGATAG 1140  
TGGCTCATTT GGTCTCTGAC AATCCAGAGC TGGGGGCTTAA AAGGCTGTGA CAGTTAAATTC 1200  
CCCTTCCCTG CATTAAACT TGGAGAGGA GACTGACAGG CTGGCCGAC ACAGGCTCTC 1260  
GTGGCCTTTT TCTTCAGACA GAGAGGTGA GCAATTAAT CACTGTGAC CAGATCTAC 1320  
TCTTCACCC CAGACATGA AACATGACT TTCTGGACAA AAGTCAAGGT GGGCGGGGCG 1380  
CTGGAAATAC AACCTGTGA GACTGCTTAC TCAACTGTG TTCTTAATTA AATGTACAGA 1440  
GAAACCAAA AAAAAAAAAA AAAAATGGA GGGGGGCGCG TA 1482

50 (2) INFORMATION FOR SEQ ID NO: 251:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 834 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

60 GCGACAGAG CCGTTCGCG CTGGCCCTT ACGATCTCT TACGAGAT GAAAGCTTT 60

GTGAACCTCT ACCAGAGAT GATGAGACAC GCAATATCCC GATATCGAGG CTACCTCTTG 120  
ATGGGATCCC CATTCTTAAT GACTTCATTT CTCTGACATP AACTGTAACTT CATTCTCTCA 180  
CTTGGGCTCT GCAATGATGC TAAATCGAAG CCGTTCAGAC TCGTGGCTT CATTGATTTCT 240  
TAAACTTCT CACTGTGAG AATTCATCCC TAACTTGTCT ATGATTTCTT GATTTCTGGGC 300  
TGGTGAAGA CTGAATCTG GCGCTGTGAC CTTGAGACAT GCACTTTTGA GAGATGTCCG 360  
TGAATGCCCT CTCCAGTAC ACCTGTGAC AAGGCTTATG TTGTGACAGC TGAAGAGGGG 420  
ACTGAGCTCT CCGTCCGAC AACCTCTTGA CAGGAGCAC GACTTTGGTT CCTGACGAC 480  
TTCCCCGGG CAGTCCAG GATGGGCTT CATTGCTCTC TCCACTTCA GAGCTGGGGG 540  
CTAAAGGCG TGTACATTTA TTTCCTCTCT CTTCCCTTAA AACTTGGAG AAGAGCATCT 600  
AAGGCTGGCC CCAAAAGAG TTCTGTGGCC TTTTTCTCA CAGAAAGAG GTGAGCAATTA 660  
ATGTCACTCT GAGCCATCT TCACTCTCCC AACCCAGCA CTGAGCAAT ACCTTCTGGG 720  
CGAAAGTCA GGTGGGGCGG GGGCTTGGGA ATTACAGCTG TGAAGCTGCG TTATCTCACT 780  
TGTGTCTTAA TTAAATGTGA CAGAGAAAC CAGAAAAA AAAAAAAAAA AAAA 834

30 (2) INFORMATION FOR SEQ ID NO: 254:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1508 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

TTGAACCTTT AAAATTTTGA ATCCAGAAC TCTAAGATCC TGAATGTGA GCTGTCTCTC 60  
ATTCTTCAT GCTACAGCTC CAGATGAGC GAGATGTGA AAGATCTGA CCGGCGACAA 120  
CGACGACAT TGGAGTGA CTTCAATCAT GAGCTTGTGAG AAGACTTTGAG CAGATGTCCC 180  
AAGTGTCTG GTGGCAGAG CTTTGTCTTT GTCTGCAAGA CTGTCAATGA GATTTGATCTC 240  
CTTCCATGAG AACAGTTTGC TTGTCACTCT ATGGCCGATC TGTAACTTT ACGAAATGAC 300  
AAGTTTCTTA AACTGCAAT GCTCTTCTCA AAGACATTTA GACAACTCT ACTTAAAAAA 360  
GACTTAATCT TGGCTCTGC CAGCTCTCAC CAGAGAGCTG TGAAGACAC CATTCAATGCT 420  
CTTCAATGAG AACGTAGAG GATGTACAG TATTTTGA GATATCACC TGCAGATTAC 480  
AAAAATCCC AAGATGCCAT GAGCAGAGG TCTTCATCT ACTTAAAGCC TTGAATTTGAG 540  
GTGTCTTCC TTCTGTCAAG AAGAGGAGG TTCAATGGAG ATTGACAGC CATGTAACT 600

GGGATAGCTT TGGGGGAGG AGAGACCTTC CTCCTCTGG GACTTCATNG CAGGTGCAG 660  
TTGCTTACAC CCATACACG GATTTTCAG ATGTCAGAGA AAGTACAGTA AACACTATTA 720  
5 TCTTATCTTG ACTTTAAGG GAAATATTT CTCAGAGAT TATATTTCT ACCGAGCCT 780  
TAAATCTTC TGTCTCTCT ACTGAATGA ACTGATNG GCAGACANT TTCTTTATGG 840  
AAGGATGAG ATTTCAGAG ACTGCATNG CTTCCTCTG GTTTTATTTA ACAAATGACA 900  
AATGAATTC TTACAGCTTG AAGCCAGAG TGTGCCAGG TGTGAAGAG ACTTCAGTA 960  
TGAGCCCTTA CTCCTCTCT CAGGAGGA CTTCCTGGG CTCGTGGCA GCTGTCCAG 1020  
10 CGAGCCCTGT GTGTGAATG TTTGTAGCT GTCCAAATG GAAAGAGG GTTTTACNT 1080  
CTCTTAAGG ACTGTATGC AACACAGTA GAATGACTT AAACCTTTA CCGAGCATA 1140  
TTGCTTACA CATTTACAG ATGTTCTCT AGTGTCTGT TGTGATTTT TCATGCTGT 1200  
20 CATGACCTGA AGGAATTTA TTAGACGTAT AATGTATGTC TGTGTATTT AACTTGATCA 1260  
TGATCAGTTC TGAGGTGCA CTCCTTACA TACTGTACAT ACCTGTAGC ACTCTTGGG 1320  
25 GTGCTGAGT CTTTATCAT GCTGTTAA CTGTGTGGC ACAATCTTC TTGTCCAAAT 1380  
AAATTTTAT AATAGACTT ATAGAGAGG AATATACAC TTTTGAATG TTTTACATG 1440  
TCTACATA AATCAATTT GTGACTGTA TTAAAAAAA NTAAAAAAC TGAGGGGGG 1500  
CCGGTAC 1508

(2) INFORMATION FOR SEQ ID NO: 255:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2514 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(11) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

GAGAGACTCA CACTCTCTTT CCATATAC TCAGCATGTA GTGAGATG CAGGGAGG 60  
GCACCTACT GTGTGTGTA GGTTTGTGA TGAATCTCAT AACCTAGAG AGGATTTAT 120  
50 AGCTTCTG CTTATGAG CGAGTCGGA AATTTGGCT GTGAATTC ACATATGAT 180  
AATGAGAG TGGGATTTA ATATGAGTA TTGTGTGGC CAGCTTACA TTGCTCTAG 240  
TGGATTTCT TCCAAATGA AAGTTGTGC TTCTAGACTT TTAGGAAT ATGCCAGC 300  
55 TATCTACCA CTCCTCTT CTTGTGCTT AATATGTG TTGGCAAT CAGTACCT 360  
TATGAGTA TCTGTGCT TAGGACAT TGAGGAAT TGTCTTTT TCCATGATC 420  
60 ACCACACTG CTTTATGAC TTGACAGT AATTTCTGTT CTTTTCAG ACGTAAAG 480

AAGCGTAAA GACTGAGG AANTCTCCA TTCTCATGTC ACAGCAGGC ATGATCTTT 540  
TGAATTTTA GTGCACTCC TCGAGCACT TGTTTTATG TTAGATGTA TAAATAGTA 600  
5 CACAAATTT AGTGGGATA ACTATATAG TGGCGAGCA TTTGTACTCT CGATGCACT 660  
GTGAGTTTT GATTTCAATG TTACTATGTT TGTCTTTAAA AATGCTCTAT CTTTTCAGG 720  
AGCTTTTGG AAAAACTCC AGGGCAAC CTCGTATGTC TTCTTTGCG CCGTAGCTT 780  
GACTGAGTA CTGATTCAC TCAACAGAT GATTCGAAA TATTCAGAT TATTCAGAT 840  
TTTGTGTGA GAGGCCACA AATTTGCCA CCAACTTCA TATTCUANT AAACCTCTG 900  
GGAATTCGG CAGGCTCAC CAGGTAAT TCGAATCTCA CTTAACTCT GAGGTACT 960  
ATAAGAAC CTTAGTCTC CCAAGTGG AGCAATAT TCAGGAAT TCAGGAAT 1020  
20 TCTCAGACA GCACCTCAA GCTCTTAAT GCTTATCTCT GGTACCTCA GTCTGGGAC 1080  
AATCAATTT CATACGTGG GAGGAACAC ATGCTGACAT GTATAGAGT GACTTACCA 1140  
ATCTGACAC GCTGTAGCT GAGCTCAAT GTTGGAAAT CAATGAAA CACAGGGGA 1200  
25 AAGTATAGA CTTTCTGTC ACATCTATG AAGCCCTCA CCTGCTGAC ATCAAGTTT 1260  
TTCTTAATG GTATGATG CTGAGGTGC TGTGTATCT TCCTGTGATG AGGTGAGA 1320  
30 ATGAGGCTA TGAATGGA CGAAGCTTC TTAAAGATA TTGGAAGAC ACTTTGACG 1380  
ACCAAGCTC AATTAACCTG CTTTGTCTTA ACATTAATTT TGTATTAAC CAGGACCTG 1440  
ATTAAATGT GACACATAT ATTAACTCT ATACAGTAA GTCAGAGCTT CTTACAGTA 1500  
35 ATTCCGAAC TGTGGAAAT ACTTAGAGA CTTTAAAA TAGCTTTCT TATATTTGAT 1560  
ATTGGAAGA AAGCCCTTA AGTGTATGTA GACCTTAA TCATTAATA TCTTTGCTTA 1620  
TAGGACTCA TTGATACAT TACCATTTG TATCTTACT GTTTAATGG CCGCTGTTG 1680  
AATCTCAG CTTTGAAGC CTACCTGTC TTCCAGAGA GAAGTTGAA AGTGCATGT 1740  
TTCTTTTGC GTATCTCTG TTGATGGAC TGTGAAATG TTTCAGTTA GTCATTTTAT 1800  
ACATAGCTT TATATCACT GTGATCTCT ACTGTGTCG TGTATGAA TCTTTGAGA 1860  
AATATATTT GAAGAGTGT GCGAGAGG AATACATTTT AATAATGTT TAGTGAGC 1920  
50 CCACATTTA CTTTGAATA ATAGAGTTT TAAGTATGTT AAAAACTAT ACTGACAGT 1980  
TACAGAAAT TACCGAGAA AACTTTGTA CTTCAACAA CAGGATTC AGTGTAGAT 2040  
TTGCTTTCT TGAATTTAA GAACAAATG ACAAAGTTG AATGGAAG CCGCTGTTG 2100  
55 TTCCACATCT CTTCTCTT TACATCTCT TGTGGGCT ACATCTCTT AAGCTTTT 2160  
GAGGTATAT GTTGAACAT TCTGTTTCA GTTGAAGCA GATCAGAG CCACTGATAC 2220  
60 TGACACTGA TTTGCTGTT TTTTCTCTT GTCTTTTCC ATGACTCTTA TATATGCTT 2280

CACTCTGAT TATACGAAA ACTCGAAA CTACAAAT AACTGTGTG GTTATCTG 2340  
AAAAATTCG AAAATTCGC TGTATTTTT GGTAGAAA ATCAATTTG TATATTTAT 2400  
TTCATCTGA ATAAATGTG AATTTGTTT AATCTTAGG CACTATATT TTTGTGGGT 2460  
CAAAATTCG TTTGTAAAT TCTCTAAC ATTGAATAA CACTTCGA ATTC 2514

## (2) INFORMATION FOR SEQ ID NO: 256:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2357 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

CTGCTTATG AACCGATCC AGAATTTTG GCTGTAAAT TTCACACTAT CATATCTAG 60  
AAGTGGGAT TAAATTTGA GTATGTCTT GGCAGACTT ACATGTCTC TATGTGATT 120  
TCTTCGAAA TGAATGTTT TCTCTTACA GTTTTAGAG AATATCCCA AGCATCTAC 180  
ACACTCTCT GTTCTGTGC CTAAATATG TGTGTGCAA AATCTATCC TGTATGCGA 240  
GTATCTGTG CATTCAGAC AATGAGAGA GTTGTCTTT TTTTCATG ATCCAGCAA 300  
CTGCTTTAG AACTTCACA CCAATTTCT GTTCTTTTG AGACAGTAA AGAAAGGCT 360  
AAGAACTGA AGGAATCTG CCAATCTAG TGAACAGCA GCGATGATC TTTGAAATT 420  
TTATGTGAC TCTTCAGAC ACTGTGTTA TGTTCATG GTATTAATG TGAAGCAAT 480  
ATTATGCGA AAACTATAT AACTGCGCA GCAATCTAC TCTGAGTGC AGTGTCAAT 540  
TTTGATTTCA TTTATCTAT TGTGTGTTT AAAATGTCC TATCTTTAC AAGAGCTTT 600  
GGGAAAAAC TCGAGGGCA AACCTGAT GTCTCTTTG CCGCCGTTG CTGACTGCA 660  
GTATCTCAT CACTACAGA AGTGAATGA AATTTTGA GTTTATCAT AATTTTGTT 720  
TGAAGAAC AGAATTTTG CACCAACT TGAATTTCA ATGAATCC CTGGAAATT 780  
CGGAGACT CACAGGGTA ACTTGATC TCACTAAC TCTGAGATT ACTTAAGA 840  
AAGCTAAT GTCCAGAG TGAACACT TATTCAGAA CTTAAGATA TATTCAGCA 900  
AAGAGCTC AAGCTTTA AATGTATC TCTGTACC TCAATCATG GACACTGA 960  
ATTCAATAG TCGAGGAG AACATCTGA CATATTAAG AGTCACTAC CCAATCTGA 1020  
CACTCTGCA GCTAGCTTC ATTGTGAG AATCAATG AAAAGAGG GAAAGAAAT 1080  
AAGCTTGG TCAACATCT ATGAAGCTT CCACTGCTT GACATCAAT TTTTCTTAA 1140

TGTATGCA TGTCTAGG TCTGTAT TCTTCTGTG ATGAAGTTC AGAATGAGG 1200  
GTATGAAT GAGCAAAC GTCTTAAC AATTTGAG AACCTTTGA CAGACAAAG 1260  
GTCAATGAC TTTGCTTTC TTACATAAA TTTGATATA AAACAGAC TGAATTAAT 1320  
GCTGACACA TATATTAAC TGTATCAG TAAATCAGG CTTCATAG ATAAATCCA 1380  
AACTGCGA AATCTTAG AACCTTTA AATAGCTT TGTAAATT GATATTCGA 1440  
AGAAAAACC GTATGTAT GTAGACACT TATATCTAA AATCTTGC CTATGAGACT 1500  
CCATGATA CATAGCCAT TGAATTTTA CTTGTTTAA TGGCCCTGT TTGACTTC 1560  
AACCTTGA GACTTACT TTTTCGAA AGAAGAGTT GAAATGCCA TGTTCCTTT 1620  
TGTGTATCT CTGTATAGG CACTCGAA TTTTTCAGT TAACTCATTT TAACTATAC 1680  
ATTATATC ACTGTGATC TCACTGTT GCGTGTATG AATCTTTGA AGAATTAAT 1740  
TTTGAGAG TGTGAGGA AGAATACAT TTTAAAT GTTGTATGA ACCCGCAAT 1800  
TAACTTTGA CTAAAGAG TTTTAATAT GTTAATAT TATCTGAC ACTTCAGAA 1860  
AATTCAGCA GAAAGCTG TGAAGTACC AAACAGAT TTCAGTATG AATTTGTT 1920  
TCTGAACT AAAGAAACA ATGCAAACT TTAATGGA AACCTGCTG TGTTCGAA 1980  
TGTCTGCT GTTATATC CTGTGTAG CCAATCTT CCAATCTT TTACAGCTA 2040  
TATGTGAC ACTGTGTTT CATGTGAG ACAATGAG AAGCATGTA TACTGACAC 2100  
TCAATCTCT GTTGTGTT TCTGTCTTT TCGATGACT TTAATATCT CCAATCTG 2160  
TAAATTAAT GTGAATTTG TTTAACTT AGCAGATTA TTTTGTGAG GGTCAAAA 2240  
TCTGTGTA AATCTC 2357

## (2) INFORMATION FOR SEQ ID NO: 257:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 689 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

ACTTCTGCT GCAAAAAGT GTTCAGCTT TATTTTAC TTGCTGCC CTTCCTTT 60  
CATTAATG AGTGAAGTC AGCTTAAAG AAGCTGTTT TTTGAATG AAGATACAT 120  
CAGAACAG GATGTGATG CAGAGTGC TCTGCTGCT TCGAATATC TCAACCGGG 180

5 ACCTCTTACT GGCAGAGACT AGAGATATG TGCACGTACT TCCATCTCTC TCTCTCTCTC 240  
CGATTTTAGC CCAGCACCAC AGGTTAGGTT CGAGTTTTC TCTCTTTTCA TAGCTGTAG 300  
GCCCTTTCTG GGAATGGTTC TCATCTCTCT TAATCTATTIA TTGGGTCACT TTCTCTGCAT 360  
GTCTCCAGCC TCCCATCACT GCCACCACT CCCACAGAG ATGCCCTCTCT CATTCGACTG 420  
GGCTTTTGAC TCCCACTGTC TGTACCCCTC TTGTGTGGAC GGCCTCTCTCC CAATACTTTC 480  
AGCAACAGC TTTCATATG GAAGTTGTCA CTGTCAAGCS CTTTACATTC AGCAACAGCA 540  
AATCTACAT GTCTCTGAGG GTCTCTGCTC ATTAGATGC AATTAATATG TAAATACATA 600  
AATAACAGCA TAGAGAGAAC GTAACTCTTT ATTCTCAAT ATGATGTCTT ACHTAGAATA 660  
GCCAATATA TTAGATAG TAGCAATT 689

## (2) INFORMATION FOR SEQ ID NO: 258:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2377 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

30 TGCACCCAGC GCTCCGCCA TGTGATGATT CTGCGTATT CTAGAGAGCC GGCCTATGCC 60  
ATCTTCTTCA TAGTCTTTCAC TGTGATAGGG GAGGCCCTCCG GCGCTGTCTCT ATCCCTGTCC 120  
GGCCACCTTT GCGTTGGTATT TGCCTCTGTA CTGGTGGGCC CCTTGACGCT GCGCTGTCTC 180  
TCTTCAAGCA ACCCTGTTTC TGATCAAGCT GCTGACAGCC ATCATCTTCA GTCACTTCCG 240  
GGCTTACCTG ATGAATCTC TCCAGACCTC GCTGTTTCCG AGCGGGTGG GACCCCGGCT 300  
GCGTTTGAAG TCTTATCTTC CATGTGCGG GAGGAGAGAG CATTTCCTCA GGCAGTTGGG 360  
GTGAGCCCC AGAATCTCT GAGGTCTCTT CAGAGGTCC AGCTGAGAG CTCCCHAGA 420  
CAGCCCATGA TCGAGAGGT GCGTTCTTAT GGCAGTGTTC TGTCTCTCAGC TCGAGATTT 480  
CAGAGCTCT TCAAGAGCT TGACAGAGT GTGGTTAAG AGCACCCTCC GAGGCCCTAG 540  
TACCACTTC CTTTCTTCA GAGGACCA GTTCTCTCTC GGCAGTAC TACTTTGACT 600  
ACCTGGGGA CCTCATGCC CTGGCAACC TGGTGTCCAT TTGGGTGTTC CTGTGCTGG 660  
ATGACAGAT GTCTGCTCTC TGAGGTGAT GACTTCATCC TGGGGGTCT CACTGCTTC 720  
TTCAATGTGT ACTAGCTGTT GAGATGCTG GCTCAAGTTC TTTTGGCTTG GCGCTCTCCA 780  
GGGTACTTGT CCTAACCCCA RGAATGTGTT TTGAGCGGCC TCTCTCAAGT TTGTCTGCG 840

5 TGGWKGSM GATCTCAACT CTGCTGTGT ACCGATTTGCC ACACCCAGCC TGGAGGCCGG 900  
AATATGTGG CCTCTGTGCG CTGTGGACA TGACCCGCAAT ACTCAAGATG CTCACTGTGT 960  
TCCGCTTCTT CCGTATCATC CCGAGCTGA AGCCGATGGC CGTGTGGGCC AGTACCGTCC 1020  
TGGGCTGTGT GCAAAACATG COTGCGTTTG GCGGATCTCT GGTGTGCTC TACTAGTAT 1080  
TTGCCATCAT TGGGATCAAC TTGTTTAGAG GCGTCATTGT GCGCTTCTCT GGAACAGACA 1140  
GCGTGGCCC TGCAGATAGG TGGGCGCCTT GTGGAGCTTT CAGAGACCTG GATTACTGCG 1200  
CCACACTT CATTGACTTT GCGGCTGCC TGTCTACTCT GTGGACTTGT ATGCTGTGTA 1260  
ACAATGSCA GGTGTTTCTG GATCTATATC GCGCTACTTA AGGCCCTGTG TCCAGATCT 1320  
ATTTGTATTT GTGTGGCTG GTGTCTCTG TCATCTGGGT CAACCTGTTT CTGGCCCTGA 1380  
TTCTGGAGA CTCTCTTACC AATGTGGACC CCGCAGCCA CCGTGCAGCC CTTCCTGGGA 1440  
CCCCAGAGC CAGCTACCG ATGACTGTGG AGCTCTCTGT CAGGATATTT CTGAGAGAGC 1500  
CCGGGAGGA TGAAGTACA GAGAGGCTGA GGCAGACCC GACCTGTGG CTGTGAGGT 1560  
GAGCTCCGG TCTCCATCC CAGCAGGGC GCGAGAGAG AGAGGCTGGC ATTAACACAG 1620  
TCCCATCAT GAGAGAGCG GCGATCTGT GCGCAGCCAG GCGAGAGAG ACCTTCTCTC 1680  
TGACGACCA CTAAAGTGG GACAGGACC AAGTCTCTTG CTTGTGGGCC AACAACATTT 1740  
TACAGACAG CTGCTGTGTC TTGAGGAGG GCGCTGTGCC TCGCTTCTT TTTATAGCTG 1800  
CTTCACTGAG AATTCCTTG TCGACTCCAC AGGGAGCTTT CAGACAAAAA TCGAAGAGC 1860  
AGCGGCTCC CTTGTCCCTT CGAGCTTCCG TGTGTCTCTT GCTGCGGCCA GCGCTTGGGG 1920  
ACACAGGCC TGACAGGGC CTGACAGGT TACCGCTCAG ACTTCCGGGG CATTCAGCTG 1980  
GGATGTATC TATATCTCC GATTTTACC CAGCACACA GGTAGCTTTC CAGTTTATAT 2040  
TTCTTTCAT AGCTGTAGG CCGTTTCTG GATGTGTAT CATCTCTCTT AATCTATAT 2100  
TGGTCAATTT TTCTGCAATG TCCCGAGCT CCGATCACTG CACCCGACTC CCGACAGAGA 2160  
TGGCTCTCT ATCGACTGCG GGTTTGACT CCGACACTGT GTACCCCTCT TGTGTGAGG 2220  
CGCTGCTCC AAAACCTTCA GCAACAGCT TTCCAAATGG AAGTTGTAC TGTGAGGCC 2280  
TTTACATCA GCAACAGCAA AATCTACATG CTGTGAGGG TCTGTGCTCA TTAAATGCA 2340  
ATAATATGT AAGTACATTA AAAAAAAAAA AAAAAA 2377

## (2) INFORMATION FOR SEQ ID NO: 259:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1193 base pairs  
(B) TYPE: nucleic acid



(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

5 TGTGATGCCC GTGCCCCCCC CCGTGGCTTT TGCCCCGCTG GCGGGAAGCTT CATTGTGCTG 60  
ATTTCGAAG ACTCCAAAGC GAGGCCGGGG ACTGAGAGTG TGGGTGTGCA GCGCTGTGGC 120  
AGAGGGTTAA CCGGGGTCAA ATGCACGGAT TCTCACTGCG TACAGTTACG CTCGCCGGG 180  
GCACGTCCGC GAGGATTGTA AGTCCGTAGC GCTCAAGTTT GTCCGTAGTC GAGGAAAGGC 240  
CATGAGGTG CCGCAGCCCG CACCGCGGAG CTTCCTCTGT AGAGCAATTG GCGTAATTTCC 300  
CGGAGTCTTT GCTGCGGAAAG CTGTGACTGC CGATTGCGAA GTCCCTTGAG AGCTGCGAAA 360  
GCGCTTCCG TACTGTCGAG AGCCCTATTG CCGGGAATCT GAGTGGAGCC GCGTCCCGGA 420  
GCGTTTGGC AAGGACAGAG TCGACACTAG TGTGAATGTA TACCGAATTA AGAGTCCCTT 480  
AAGCCATTTT GTTAATTCAG GAGCTGTAC GCGAAGTCTT TTTAGATTA AGCTAGCCCT 540  
GCGTGGCTGG TGGCTGGTGG CAAATTTGGA GCGTTGCTGG GAGTCTCTGT AGAGGCGCTG 600  
CTGATGGCAT TTGAGAGTA CTCTGTGAG ACTGTTCGAG AAGAGAAACA GAGGAGTGA 660  
AAGCAGTCC ATGAGCTAAA ACTGAGAGAG TCGAAGGCA GACTGAGAT TACTGAGCAC 720  
CTCCCTGAGA AATTTGAAAG TACGTTACAG GAGGATGAGC CTGAGAAACA TCGTAAGAAA 780  
ATTGAGCAG TCGTAAGCT TCGTAGAAG CTTTCACTTA TTAGTAACA AGACAGGAGC 840  
TGAAGTGTCT CTGAAGTGA AACTGACTGG AGAGCTGAG GAGCTGCCA TGTCCGATGA 900  
ATGCCAGAG AGAGCCGACT CTGTGTGAG CCGTGTGAGA AATTGAGTG CTGTGAGCTG 960  
TGTGTGAGAT GCGTGGCTCT TGTCTTTTTC TTTTCTTTT TACTAGGAT GCGGCTGTTC 1020  
TACTGTGACT TTAATTATCC TTAATTTA ATACATGACT ATGTTGTAT TAACTATCA 1080  
ATATATGAT AATATATAT ATGCAGCCAG CTGATTTTA AGCAGTAAT AATACATTC 1140  
GCAAAAGATT AAGTTGAT TTTACAGTTA AAAAAAAAA AAAAAAAAA AAA 1193

(2) INFORMATION FOR SEQ ID NO: 260:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1262 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

60 GAAAAAGCA AAGATCCAGA CAATCTCTTT GAGCATGAT TGGGAGCTCT CAATATGCT 60

GCATTACTAC GAAAAAGAA AAGACAGAT CTCTTAGTA ATCTGCCC ATGTGTAGC 120  
GCGTGTCT TCGAGCCGA TTCTGCAATT CGAAGCAGC TTGTAAAA TGAAGAGCC 180

5 AACCATAAC AAGCTACAG GATGTCTCCA ATGTAGACA ATGATTAAT TCGATTGAT 240  
CTTGCGATT TTAAGCGAA GACTACTTGC CATCTCCAG GAGTGAAGA GACTGAAGAT 300  
ATATACTTT CAGATCCAG TATCCAGAG GAACTGTG TGCTTAGA CCGAAGTTCA 360  
ATATGATAT TTTGGTACAG TGTCTCTCA GCAATGATA TCTTTTCA AAGTCTTTC 420  
GTTCAGCAG CATTAATGAC AAGGACAGA AAGATTAT ACCATGCTA AAGAGTGA 480  
GAATTTGAT CTTHAGAGC ACTAGTTTG GCAACTTAA GATTTAAGT TAAATTTTAC 540  
ATAGATTTG ACATGATGC AAAATATAT GAAACATCT AGATTAGTA GTTAATCTG 600  
GCGCTTTGT TAAACTGAA GATTTGAA ATGTGTGTC ACTGCTTTC GAGCTTAGA 660  
ATATTTTGT GAAATGAGC CATGATTA TGTGTGATC ATCCATAG AAGCAAGAT 720  
TTTATGAAA AAGATATGT TCAATCAAGT AATTTCTAC ATGTGCAAT ACTATGTGT 780  
ACAGACAG TGAAGGAAA TGCTGTGTA GATGCGTGG TATGTATTA GCGCAATTC 840  
ACGAGAGAA AGCAAAATA GTTTTTCCT TTGAAGATT TTTTAAAT TATTTGATGG 900  
GTCTTTTAT TAAATATAT GTTGCAAT TTAGATTA TGTTGATGT CTTTGAGCC 960  
TAAATGCTT TTTGTATC AGAATGTG TACTATTTT AATTATTA AATGATCTT 1020  
CCCTTTCCT GTTTATAT TACTTCTC TGCCTTATC TTATCTGTA TGAATGAAA 1080  
CATATGAT CAATATCA TGTTCATGC TTGAGATAT TTAAGCTTG TGTCTTAT 1140  
GTTCAGCAG TTAAACGA GTTAGATGA CTTCATAT ATCTCATGA TACTTAAGG 1200  
CAATCTCCT GGAATGAGG CTTTTGAA GAAAAAAT TACCCAGAG GCAATCCCA 1260  
GT 1262

(2) INFORMATION FOR SEQ ID NO: 261:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1179 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

55 GCGAATCTT CCGCAGAGC TTGGAACTT GCAAGCGAA AGCTGAATC GTTAAGACT 60  
GCGTGCAGC GCGCCCTGG CCGAGAGAG CCGAATTGC GTTCCGAAA CTTTGCCCTT 120  
CAAGGCTGG GCAAGGAGC ATGTCTTCA CCGAGAGAG CCGCCTTGG GTACAGGAC 180

CTGGACCTCA TCTTCTCG CGGACTCG CGGGAGGG GAGTTTCAGA CTGTGAGGA 240  
CGTGTGCTG GACTGCTCTT TGGACTTCTT ACCCGAGGG GTGACACAG AGAGCTAC 300  
ACCACTCAG CTCAGGAG CTATATGCA GAAATGGT AAAGTGTGCA ATGACTCTGA 360  
CCGATGAGT CTATATCCC TGTCAACA CAGTGGCAA AATGTGGAC TGAATTTGT 420  
GGATTCCTC CGGAGCGAGT TTGANTCAG TGTAGTTCT TTTCAAATCA AATTAGACT 480  
TCTTCTGTC TTTATGAAT GTTCAGAA CCCAATGACT GAGCATTTT ACCCCAGAT 540  
AATCGGGAG AGGCTTATG GCGATTCCA GGAAGCTTT GATCAGCTT GTACAGAT 600  
CAATCCAC AGGACCCAG AGGAATCG AGGGGAGGC CTGCTTAGT ACTGCACT 660  
CTTGTGAGG GCTTTAGC CCGCTCTGA TGAATCAG ACCCTTCAA GGTATATGT 720  
TTCCAGGTTT TTATCGACT TCTCAGACT TGTGAGCAG CAGAGAAAC TGGAGTCTTA 780  
TTTCCAGAC CACTTTGTG GATTCAGGA CCGCAAGTAT GAGTATCTCA TGAACCTTCA 840  
TGGATGTTA AATGAGCA CAGTGTGCT GATGGACAT GAAGAGAC AGACTTTAA 900  
CCTATCACC ATCTGGCTA TCGGTGTG AGCTGACCA AATGTCATC CTATGTGG 960  
TAATGACT TGTATTACC AGCAGACCC CTATGTAGA GATGCCACT TTACAAATTA 1020  
CTACATTGA CAGTTTACG CAGTATCAC GTGCCAGCA CAGACTACT CCACTTGGT 1080  
ACCTGCCAT TAAATATCT TTAATATGT CTTGTGGGA AGCCATTCA GACAGACAG 1140  
GAGAGAAA AAAAAAAAAA AAAAAAGC 1179

## (2) INFORMATION FOR SEQ ID NO: 262:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1162 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

GCGAAACTTT CCCCAGAC TTGAAACTT CGAGCGGCA ACCTTGATC GTTAAAGTT 60  
GGTTGGGAC GGGCCCTCG CCGAGAGAG CCGATTGGC GTTCCCGAA CTTTGGCCT 120  
CAAGGCTCG CGAGCGAGC ATGTCTGCA CCGAGAGAG CGGCTTGGC CTACAGAC 180  
CTGACCTCA TCTTCTCG CCGACTCG CCGGAGAG GAGTTTACA CTGTGAGGA 240  
CGTGTCTG GACTGCTCTT TGGACTTCTT ACCCGAGGG GTGACACAG AGAGCTAC 300  
ACCACTCAG CTCAGGAG CTATATGCA GAAATGGT AAAGTGTGCA ATGACTCTGA 360

CCGATGAGT CTATATCCC TGTCAACA CAGTGGCAA AATGTGGAAC TGAATTTCT 420  
GGATTCCTC CGGAGCGAGT TTGATTCAG TGTAGTTCT TTTCAAATCA AATTAGACT 480  
TCTTCTGTC TTTATGAAT GTTCAGAA CCCAATGACT GAGCATTTT ACCCCAGAT 540  
AATCGGGAG AGGCTTATG GCGATTCCA GGAAGCTTT GATCAGCTT GTACAGAT 600  
CAATCCAC AGGACCCAG AGGAATCG AGGGGAGGC CTGCTTAGT ACTGCACT 660  
CTTGTGAGG GCTTTAGC CCGCTCTGA TGAATCAG ACCCTTCAA GGTATATGT 720  
TTCCAGGTTT TTATCGACT TCTCAGACT TGTGAGCAG CAGAGAAAC TGGAGTCTTA 780  
TTTCCAGAC CACTTTGTG GATTCAGGA CCGCAAGTAT GAGTATCTCA TGAACCTTCA 840  
TGGATGTTA AATGAGCA CAGTGTGCT GATGGACAT GAAGAGAC AGACTTTAA 900  
CCTATCACC ATCTGGCTA TCGGTGTG AGCTGACCA AATGTCATC CTATGTGG 960  
TAATGACT TGTATTACC AGCAGACCC CTATGTAGA GATGCCACT TTACAAATTA 1020  
CTACATTGA CAGTTTACG CAGTATCAC GTGCCAGCA CAGACTACT CCACTTGGT 1080  
ACCTGCCAT TAAATATCT TTAATATGT CTTGTGGGA AGCCATTCA GACAGACAG 1140  
GAGAGAAA AAAAAAAAAA AG 1162

## (2) INFORMATION FOR SEQ ID NO: 263:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 735 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

CGGCTGGGT ATTTGGCTCG CACCATGGCG CCCAGGGCA AGTGGGAC GAGAGGGAG 60  
AAGCAGTAT TTGAGAGAA CAGAGACT CTGAAGTTCT ACCTGGGAT CATACTGGG 120  
GCGATGCA TTTACTGCT TGTAGGTTG GTCTTCTTTT ACTGATCTG CTGATTTGG 180  
GCTTGTGG CTTTGGCTT TAGTCTGCA GTGTATGGG CAGCTACCA CTCTATGAG 240  
TGTATGGAC GAGCAGGTT CTCTGAGGA TGGGGCCCTG ATGATGGTG GCAGGAGTC 300  
AAGATGAGC AGGGATGGC AGAGACTT AAGATGTGA TCTTACTGAC AGCATGTGT 360  
CAGGTCTCA GCTCTTCTC TCTTATGTC TGTGCTCTT GTGCTCTGG TCCAGCGCG 420  
GCGCTTACC TCTGTGGT GATGTGCTG GCGCCCTGGT TCACTGCGA CAGTGGCAC 480  
CGAGCAGC AGCAATCA GAAGGGCAG CCGGACAGG AGCGCGGCA GATGAAGCG 540  
TTATAGCAT TGCATTTGT GCCAGAGGC ACTGGCCCTG GTTGGCTCTG TCAGGCTCA 600

CAAGCCCTCA TCCCTGAGC AATGAGGTC TACATCAGG GCGAAGCA CTCAGCTA 660  
TTGGGTATAC TTAATCTTA TAGGTCGTT GAATTAATG CTTAAATAT GAAAAAATA 720  
AAAAAAAAA ATTTT 735

10 (2) INFORMATION FOR SEQ ID NO: 264:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 783 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

AATGATAGA GCTCCGATG TGTGCTTAG TGAATGGGT TTCCGTCCT CTCGCTGTT 60  
TCCCGGCTG GGTATTGCT TCGACATG GCGCCAGG CCAATGAG CACAGAGAG 120  
AAGAACAGA TATTGAGAG GAAACAGAG ACTGAAAT TCTACTGAG GATGATATG 180  
GGGACCAATG CCAATTACT CCTTGAGAG TTGATCTCT TTTACTGAT TCCCTCATTT 240  
TGGGCTGCT TGGCTGGGC TTAACTCTG CAGTGAATG GCGACATAC CACTCTAGA 300  
GCTGATAGC AGGAGCAGG TTCTCTGAG ATGGGACCT GATGATGAT GGCATGAGC 360  
TCACATAGG GAGAGGATG GCAAGATGAG TGTCCCCAC CCGAGCCCA GCGACCTTA 420  
GCATGATAC CTAATGAGC CATTGTGCA GATGCTGAG TGCCTCTCT TCTATGCTG 480  
GTCCCTCTG CTCTGACTC CAGGCCGAGC CATTACTCT CTGAGGTA ATGCTCTGG 540  
CCCTGCTTC ACTGAGACA GTGCGACCC AGCAGCAGG CACATAGAG AACGGGAGG 600  
CCGAGAGAG CCGGCGGAGA TGAACGGTT ATGACCATG ACATTTTAC SAGNRCCAC 660  
TGGCCCTGG TGCCTCTCT AGGATGACA GCGCTCTAG CTTGAGCAA TGAAGCTCA 720  
GTGAGAGGC GAAAGAGAT CTGAGATAT GGTAAATCT AATCTTAA GGTCTCTGA 780  
ATA 783

50 (2) INFORMATION FOR SEQ ID NO: 265:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1638 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

GCGACAGGC GCGCGCAGG GTGGGCTGG CCGCCCCCG CCGAGACCTT ACCCTTCCC 60  
GTCCGAGAGC GCGGCGCTGG GCTCCAGGG ANCCCGGAC AGCGAGAGT GAAAGAGAT 120  
GATTCGCGC GCGCTCCCC CCGATGAA GAGAGAGAA GTATCCAA AATTCGCT 180  
AATGCTGGC AAGAGGATG TCTACTACT CATGCAAT GGTAAAGT TCGAGAGAA 240  
GCTCAGTTG GCAAGTACC TGGAAATG TGTGATCTC AGCTTTTG ACTTGAGAC 300  
TGGAAATG TCCCTTAGA AATTACAGA GAAACAGAG AGCTGCGAA AGCTCTCT 360  
CAATCAAT AAGGTAAAC CAGCTTGA TACAGATG CCAATTAG AACAGATC 420  
AATTTCAA CAGCGGTAA CCAATGAC AATCATCTT AGTAATAA TGAATCAGA 480  
CCGACAGGA ATGATGAGC AGCAGCTCA GCTTTCTGG GAGAGAGC TACAGACT 540  
TATGATACA GATGATAGG AACAAATAT AAAAAAGAG GAATACCA AAGGTCTCA 600  
AGAGTTGAT CAGGTAGCA ATGATGAGC CTTTATCT GCTTTTCCA GTGCTTCCA 660  
CAGAGCTCT GCGCAATCA CAGGCAAT CTCGCTGCT GTGAAAGA ACCCTGCT 720  
TTGCTTAA ACATGACAC CCGTTCGA AGCTTTTAT GTGAGATG AAGCATCAG 780  
GAAACAGAG GAGGAGTAC AGCAGTAC CAGAAATG GAAAGAGC TGAATCAGA 840  
CATCTGTGG CAGCTGCTG ATACAGAG GATGATAT GAAATGACA GTGAGATGA 900  
AGCTTAAAG TATGATAGG TAACTTCCA CCACTTCCC CCAAGAGAA ATCTGTA 960  
ATTGACAAA AATTTTCCA CTGCTTTTG CCTGTAAGA AAAAAATGA CCGACACA 1020  
TAGAGCTTT TAAATGACT AACCAATCC TTTTAAAG TATTTTAT GTATATAT 1080  
AATATCAA AATCATGTT TATTGATG CTTAGACTT AAAATTAAT TTTTAAATA 1140  
TGAAGAGAA CCGTAAGAG AAGTGAAGT TTTGATCCA GATGATAT ACTGAAATG 1200  
TAGCATTAC GTAAAGAT TTTTCCCC ACAGTTTGA TGAAGAGAG TCGAGATTC 1260  
TAAATTAAT TCCCAATTA AGATTAATG GACTTGACT TATATACAA TATTTTATA 1320  
CTTATTTGA AAGGAGAGC TGTAAATCT TCCATCTCA CTGAAAGAC AATTAATGA 1380  
TTATATCAC AGATGATG GAAATCTTC TTTGAAAG CACAGAT AAGAGACCC 1440  
TGTAAAGCT TGTCTGAT TACATCAC TGTGATCCC GGCCTTTAG TTTGACATG 1500  
GAGTGGAG GAAATAGAG CATATATG CAGATGAC TATTCGCTT GCGAGCTTCT 1560  
GAGAAATCT GCTTACACA GAATTTTAA GAATTTCTG CTTAAATAT ACCTACCTG 1620  
TGTAAATTT TTTTCCC 1638

60

## (2) INFORMATION FOR SEQ ID NO: 266:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1455 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

5 CCGTGGTACT GCGATCGAGG TACCGGGTCC GGAATTCCCA GGGTGCAGCC AGCGGTCCCG 60  
TCACTTTGCCA AGCTACTCTGG GAATTAATCT TGAATCTCAGC AGTTTGTAGT TCAGAACTCG 120  
AAGATGATG CCAATGTAAT TACAGAGAGA CAACAGAGA CTCGAAAGC ATCTCTCTCA 180  
TCAAAATAG GGTAAACAG ACTTGATATC ACGNTTCCA ATTAGACAAA CAGCATCAT 240  
TTTCAACAA CCGGTAAACCA AAGTCAACAA TCATCTTACT ATTAATAGTCA AATCAGAAC 300  
ACACAGATG AATGACAGC CAGCTCAGCT TTTCTGGGAG AAGAGCTTAC AAGGACTTAG 360  
TGCATCATG GTACAGAAC AATATATAA AACATGAAA CTACCCAAAG GTCTTCAGG 420  
AGTTGGTCCA GGTAGCATG ATGAGACCT TTTATCTGCT GTTCCAGTG CTTTCCAGC 480  
AAGCTCTCG CCATATCAG GCGAATCTC CGCTCTGTG GAAAGAGCC CTCTGTGTTG 540  
GCTTACACA TCTCAAGCC TCTGCAAGC TTTTATGTC ACAGATGAG ACATCAGAA 600  
ACAGAGAGG CCAATCAGC AAGTACGCAA GAATTCGAA GAAGCACTGA TGGCAGCAT 660  
CTTGTCCGA GCTCTGATA CAGAGAGAT GATATTTGAA ATGACAGTG GAGATGAGC 720  
CTAAGATAT GATCAGTAA CTTTGCAGG ACTTTCCCA AGAGAAATTT CTTAGAAAT 780  
GAACAAAT GTTTCAGTG GCTTTGCTT GTAGAAAAA AATGTACCC GAGCAGATG 840  
AGCTTTTAA TAGCACTAAC CAATGCTTT TTAGATGAT TTTTGATGA TATATCTATT 900  
ATTCAAAAAA TCAATGTTAT TTTGATGCTT AGGACTTAA ATTAGTCTTT TGTATATCA 960  
AGCAGACCC TAAGATGAG CTGAGCTTTT GATCCAGCT GCAATCTACT GGAATGTAG 1020  
CACTTACGTA AAGCAATTTT TTCCGCCA GTTTTAATA GAACAGATCA GGAATTTCA 1080  
ATAAATTTCC CAGTTAAGA TTAATGTGAC TTCACTGTAT ATAAACATAT TTTTATACT 1140  
TATTGAAGG GGAACCTGT ACATCTTCC ATCTCAGTG TAAGACAAA TAATGATTA 1200  
TATTCACAGA CTGATGGA TTTCTTCTGT TGAAGAGAC ACACATATA GAACCCCTG 1260  
TTAGCCTTC TCTGATTAC ATTCACTCT GATCCCGGG CTTAGGTTT GACATGGAG 1320  
GTGGAGGAA GATAGGCAT ATTTTCAG TATGACTAT TGCCTCTGG ACCTGTGAG 1380  
GAATCTGCT TTCACAGAA TTTCTAAGA TTTCTGCTT AATATCAAC TAGCTGTG 1440  
TAATTTTTTT TCCCT 1455

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1003 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

## (2) INFORMATION FOR SEQ ID NO: 267:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1086 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

5 GCGCTCAGT ACCGTCCCG AATTCCTGGG TCAGCCACAG CGTCCCTGAC CCAGAGAGAG 60  
CTGCTGTCT ACATCAGCT GGGCTCAGC GCGTCCCGC GCGGGGGCG CGAGCTGAC 120  
TATGTCTCT TCAGGGCGG CACCTGTGG CATTAATCTT TGTACCCCA GCATCTACCA 180  
GTGTGGCAT GTACTAGCA CTCAGAAAT GTGTGTTGAA TGAAGATGC CTGTGACAG 240  
CAGCGCACT TTATCTTTC CTGACCTTG CTCTATGAC ACACCTCTC CTGACTGCCA 300  
CTGTACTCC TTGAGAGCAG AACTCTCTTA GGGAACTCG ATGGGAAACA GCGATGGCA 360  
AGGACATCT GGTGAGCA GCGTACACT TTGATGAAT GAACAACTG AGGCTGTGCG 420  
ACCGAGGT TACCCAGCAG ACCATGAGC TCAGGAGA GTGCAAGAC TTTGTGACCA 480  
AATTTGCCA GTTTCAGAA ATAGTTGGTG GTTAAATGA GCTTGTGAT CAACTTGC 540  
AAGAGCAGA AATGAAAG ATGAAGCCA TCGTCTCTG GAATCTCTC AATCTATAG 600  
CAAGCAGAG AAGAGCTCA CAGCAGCAC TTCAAGCCT ATAGCAGAA AAGAAATGC 660  
AGCTAGAGG GTATCGGTT GAATGTGAG CTTTGTGAA AGTAGAGCA GAACAAATG 720  
AATTTATGA CCAATTTAT TTTCAGAAT GAAGTGAATA TTTGCTTTT ATAGTAGAA 780  
GCCAAACAA AAAAAAGCT CTCAAAACA AAAAAACCTC TGTAGCATTC CAGCGCTTG 840  
ACCAATGAC TATGTCAAA GAGTGGGT GTAAGGAATG CAGCCCCCTG AAGACAGAC 900  
TACAGCTCG GGGAGCCAG TTTTACATC AGTCACAGC TGTCTCTGT GCGCTCTGAG 960  
TGTAGCTCT CAGCTCTAT GCTTAGTGG AACTAGCAG TTTGTAACT TTCATCTTT 1020  
TTTTGTAAA TTCAAAAGC TTTGGAAGA GAGCAATAA ATTTTGTCT TCNAATGCG 1080  
TTGATG 1086

(D) TOPOLOGY: Linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

5 GCGACGGGAG CAGCGGGCT GGTCTGTCTG CAGCGCGGGG GCGCGAGTGG GGGCGGGCGA 60  
GGAAACATGA AGCTTGGAGT TCCCGCAAGT GAGTGAATC GAAATACCGG TGTCAATGAC 120  
AGCGGGGGTA TGTGCTGAC AATATGATGG GAGTGGGCT TCTTCATAT TCTCTTACTC 180  
AGCATTCCTT TCTTGAGTGT TCTGTGTCT TGAATTGTA GAAATATTAAT ACATATATCTG 240  
GGATGTGAGG TAAATTTGCA TGCAGTAAA GGAACAGCTT TCGAACTCC TACCGAGGCT 300  
AAAAAGAGG CTCCTTACTC AATGGAGACA ACTGACATAT GAGATGAGT TTACATCTTC 360  
AGCGAAGTTT TTGCAGATTT CTCGATTAAT TGTATATTTT CTGCGAAGTT TGTATAGAAA 420  
GTATGATGCA ACTGACTTCA TGTATACAC AGCTTCTCTC CTGAGTGTAC TAAATCCGAA 480  
AATGCCACAA CTAGATGGTG TTGCGATCTT TCGAATTAAT AAGTATGGA ATGTATTGAA 540  
ACTGAAAAAA AATTTCAGG CTACTGATTT TCTTATAGAG AAGAGTGGT TATGTAAGTG 600  
CACTGTCTCT CTGATATGT GAAATGAAA GTATTTACAT TCGAGCGCCA ATGCCGTGTC 660  
CTTCAGATGTC TGTATTGAG TGCAGATTTT CATTAATGCA TGCCTGTATT TAAATACACT 720  
GGTACATTC TGAAGAGGG GTTATACAGC AGCGTGGCCA GCGCGAGCTT ATATATTA 780  
GGCATGCA CTGAGGTGT AGTATGATTA TTGAGGAAA TAAAGATTT GTAAAGACT 840  
AGGACAGACT TAACTATTA TGAATGGCA TTGTATTAG AAAAGACAT TTCCAGTCAAT 900  
TGCAGCTGGG TTATTTAAG CAGACTTACA TGTAAACCGG AATCTCTCT ATACAGATTT 960  
ATTAAAGATT ATTTTATTA CCGTAAAAA AAAAAAAAA AAA 1003

(2) INFORMATION FOR SEQ ID NO: 269:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1234 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: Linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

50 AATGAGCATT ACAATGACA TATTTGGAT GGTGTTGTTG TGTACTTCA AAGTATGAG 60  
GAAAAATTA TCTTCGCAAC AGAGTACTT TGTCAATGCA GAAATGGGGG TGTAGATGG 120  
CGAGTTGAT CAGTTGAT TCAATTCATT ACTTCTTCA GAGCAACAT GAGATGTGGA 180  
GTGCCACAA GTAAAGTGA TCGAATACG CGTGTCATGA ACAGCGGGGG TATGTGGCTG 240  
ACATATGAT TGGAGTTGG CTGCTTCAT ATTGTCTTAC TCAAGATTC CTCTTCAGT 300

GTCTCTGTGG CTGAGACTT AACAAATTT ATACATATC TGGGATGTA CGTATTTTTO 360  
CATGATGCA AAGACACAC TTTCGAACT CGTACACAG GTAAACAGG GCTCTTACT 420  
CATTGGAAAC AACTGACATA TGAATGACG TTACATCTT CAGCGAGTT TTTCAGATTT 480  
TTCGATTA TTCTATTTT TCTGAGAT TTCTATGACA AGTATATCC AACTCATCTC 540  
ATCTTACAA CAGCTTCTT CCGAGATGTA CTATATCCA AAATGCCA ACTACATGCT 600  
GTTCGATTT TGGAAATTA TAAATATGTA AATGTTTGA AACTGAAAA AATTTTACA 660  
GCTACTGAT TTCTTATAG GAGGAGTGG TTATGAACT GAGCTGTTTC TGTATATAG 720  
TGAATGACA AGTATTTACA TTGAGGGGCC AATGCTGCT CTTTCAATG CTGTTTGA 780  
GTGCAATTT CGATTAAATG ATGCTCTGT TTATTCACAC TGGTACATTT CTGACAGGG 840  
GCTTATAG CAGCTGGGC AGCGCCACT TATATGTTAA AGGCAATCAG ACTGAGGTG 900  
TATGATGTA ATTCAAGGA ATTAAGATT TGTAAAGAC TGAACACAC TTACTTATA 960  
ATGATGGGC ATTGTGTTA GAAAAACA TTTCAGTCA TTCACTGTG GTTATTTAAA 1020  
GCGACTTAC AATGTAACG GAATCTCTTC TATACAGTT TATTAAGAT TATTTTATT 1080  
ACGTACATA TTTCATCTT TTATGTTAG YGAGTGTGA TCTCTGTT TTATACAC 1140  
CAATGCCAC TTATAGGCT ACTTTTTGG TTTCGTGGG GTTATATG TGTATGTGTC 1200  
AATGAGCCA TTTTACAT TATTAAGTT ACG 1234

(2) INFORMATION FOR SEQ ID NO: 270:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 574 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: Linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

45 NGAATGGCT TGTAGAGCT GTTCTCTCG CAAAGATCT TCAAGATTT ATTAAAAA 60  
TATGATGCC CAGAAAGCC TACATACAA AATTTTACA GGAATTTGG AATGAAATGG 120  
GAGTATAGG GTTCAATGTT TTTAAATCC GAGCTCTGA TAAAGAGAT AAGCTTTGA 180  
AAGCTTCAAG GCTCTCTCT GGTATCACT AACCAATTT ACTGAGTGA CATTGTAAG 240  
AAAAATGAG TGTGCTGTA AATTTACAA AGCGTGTGA GATGGAGAC GTGAAAGTTC 300  
ACTTACACT TGTATATGA CCGTTTACT GATGGAGTA AATATGAT CTGTGAGATG 360  
CACTCTTAC TGTACTGCT TTCAATGTT TCCCTCTGAG CCGTCCGGG TGTACAGAT 420

ACTCTGAGTA GATTAATTTGT CATGCAGGC ATGCAATCAG AATCTACTG AGCCACCCAT 480  
CATTTTGAAA TAAATTAATCT AGTTGTACAG GACTTGTGTA TCAGATCCA GGCCTACT 540  
TGTATCTAC TCTCAATATA ACCTTTATTA AACT 574

5

10 (2) INFORMATION FOR SEQ ID NO: 271:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1731 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

20 GCTCAAGGT GCGCTCTGT GCGCTGCAGA TCCAGCTCAC TACCTTGGA AATCTTACAC 60  
CTTCAAGCAC TGTGTATTTT TCGTGTGATA TCGAGGAAG GTTCAGACCA GCCATCAAGT 120  
ATTTTGGGA TATTATTAC GTGGACAGA GATTGTCCA AGGGGCGCG ATTTTAGAA 180  
TTCTCTTTAT TGTAAACAGA CAAATACCTA AAGGTCTTGG CAGCAGGCTT CAAGAAATGG 240  
ATTAAACAG TGTAAAACGT GTACTTCCAA AGACCAAGTT TTCAATGTGA TTACAGGAAG 300  
TAGAAGCGC ATTACAGAG ATTCGCGGAG TCAGAGAGTGT TGTATATTT GAGATGAAA 360  
CTCATCTGT CATCCACAA ACTGCGCTCG AGCTAGTGTG CCGAGGAGTC GAGTTTACA 420  
TTGTCTGTA TCCCACTTCA TCAGAGACA TGATGACAG GATGTTTGCC CTCGAGGCTC 480  
TGGCTCAGC CAGGATCAT ATGACACAG AGTGAAGGCT GTTCTGCTTC AGCTGTAGC 540  
TGATAGAC CATCCAAAT TCAGGAAT TCAGATCTTA ATTAGGCA GTCTCTGGA 600  
GTGCGGTCT GTTTCCAAAG TATAGACAT TTCAAGAACT GGTATGCTAC TCAGTGTGA 660  
AGGACAGTCA GGTGAGGAC TGTAGGCCA CAGAGCTCT TCTTATCTCT ACTAGATTA 720  
AAATGTTAG TCMAAAGCG CTCCTTTTTT GCGCTCTTA GTCACTTAA CCAGCTAGAC 780  
CATTTGAGTA CCAGATTTA GTTACAAAG TCAGAGGCTT CCGGTGCTGC TTACCTTCT 840  
TTTTTTTAA TGTCTTTTA TTATTAAAA AAAATTACA TGAAGTCC TGTTTTGTCT 900  
CTACTGTGA CTCTGATCT ATCTTTCCA AGTGCAGACT CTGTGAGT TTTCTTAAT 960  
TGTCTACTT AAGAAATG AGTACACAC ATGATTTGG CTTTATAT ACTGTAGAT 1020  
GTTATATCT TAATGTGAT GTAGTGTCTT TACTTTACAG ATTGAATGA ATAGATTAAT 1080  
TGCATAGAA TTTACCAACA GACCTCTGAA TCATGTACC CACTCCCTC ACAATGTGT 1140  
CCACTTAGT AGTTGCATG ATCTATCCTT ACCAATGAT GTTCATTAAT TACTATCTT 1200  
TCTGAGAT ACTGATTTCT TATTTTGTG ACTATTAATA AATCTCTT AATATCTCT 1260

CTTTTAACTG AAGAGGATG GGTAGAGCG GTTTCCAATG CCATATTAT GTGCGAGCG 1320  
TGTTTTAACT TCTTTGAAGT ATGCTTGTCT GATATATCTT ACCACATCT TGAATATATA 1380  
TTCTAGTGTG CACAGATTT ACGAAAAGA TAAAGCTTGG GTGGAATATC ATTTTAAAT 1440  
GTTCATGTC TGTCTATAT TTCTTCACC TACTCTCAA ATATGTAAAT GCAAAAGTC 1500  
TCAGTAATGA TTGTGTAGTA TTAATTTTCT GGTCAATGTT TCTCTTCAT AAATTTATTT 1560  
TCATTAATA CTTTGTAGAG GTTGTGAAA TGTTTTCAA ATATGTGAAA TGTGAACGTG 1620  
CTGTCTTTTA TATTAAAGTA ATTAAAGAA ATGTATTTGT ATTGAATTA TTTTGNCTC 1680  
CACAAGATGG CTCATGAGT ATTCTTCCAG GATTTCTAAT ATTATTTAA G 1731

5

10

15

(2) INFORMATION FOR SEQ ID NO: 272:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1320 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

25 CTCTTTAGGA AGAGAGGTC AGAGTTGCG GGGCCAGAG CATCTCTGCC GCTGCGCCAG 60  
TCATATGTA GTGAGGGGC AGACACCCCT CCGCAAAATC TCGAAGTTC TTAGTCTGGA 120  
CTAGGGGAGT AGCCAGAGAC TCTATGTGCG CGGTTTCAG TCACTGCGCG CTGAGCGAG 180  
CTGCGCTGCG CATCTTTGCC TCTTGTGCG CGGGGAGGCT GGTGCAACA GCTGCACAGC 240  
AAGTGCAGA GCAATAATTT GTTTTGTACT TACTGTGAT TGAAGTATC AACCATGTTG 300  
TGGTTTTTAT GCTGGGACA ATCCATTTT CTGAGGGAT GGGAGGATCT GTCTACTTTT 360  
CTTATCTTGA TTCAATGGA ATGCCAGTAT GCGAACTCTT AGGATTTGTC AGGAATGGA 420  
AGCCAAGTGC CATCTTCAA ATTTCAAGTC TTAAATCTGG AGAAGGAAGC CAGATCTCTT 480  
TTGGAGCCAT GATTAATGTC CGAACTCCAT CTGTGCTCA GATTGGAAT TCAGTGGAAAT 540  
TATTAGAG TATGGCTCAG CAGACTCTTG TAGGTATGCG TCGTGTATCC TCAGTTGACT 600  
CATCTACTA GTTCACACA AAGATGTGCG ACAATTTCTA CAATTTGCT TCACTATTTG 660  
CTGTCTCTCA GCGCCAGATG ACACCAAGCC CATCTGAAT GTTCATTCGG GCAAAATGCG 720  
TTCTGCAAT GGTATGAGGC AATTTCTGTC TCAATATTA AGGCTTTTTA TACTGATA 780  
TCTATTTGT CTATGATAT ATTCTTTTTT TGACATTTAA ACAATCTCT TTATTGTGAA 840  
CATCAGACT GCATGCCAT AAGTATGTA CTATAGAGAT CTGATGAGAA ACAGTTCTTA 900

CCCTAAATAT TTCTGTATAT TGTCCCAT ATGATTTAT AAAAGACAA AATATATTT 960  
GCTATATTT TACGAGACAC TATTAAGCT TATTAATTT TGTATTTT CATTTAGAG 1020  
TACATATAT GAGATATAT TATATCCAC TGTATACAC AATGCTATT GATCTATTT 1080  
TCCAACTTC CCGTGTTTTA TGAATATTC TTTCACATT GATATCTTA GATATGAG 1140  
GATCCATAG AAGATTTAT TGAATATAG TACATATAG TACATATGA GATATATAG 1200  
AATGACATAG GTTGAATTT GCAATATAG TACATATAG ACCATGACA AATTTCTTC 1260  
ATCTGCTAT CTGATATAG AGGAAATAG AGTGAATTT AATTTTCTTA GATCTATTA 1320

## (2) INFORMATION FOR SEQ ID NO: 273:

- 20 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 515 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## 25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

CGCTGAGAG GGGCTGCTT GCCACCTTG GAGAGGTCG GATAGGGGCT GCGCTGATG 60  
GCGCTGATG GAGTACCTT GCCACACTT GCTAGGCTGA ACTTTATTT ACCCTTCCC 120  
TTGTTCTCT TATGAAAC ACAGAGGAG TGGGAGCTC AGTATATCA GAGTATATTA 180  
TTCCGACAC ACGGACTCG GAAAGGCAAT GAGGCAATTC TTTCAGAGAA TGTTCATAT 240  
TCAGCAGAA GCAATCAAT AATTAAGTC TGACTTTCTG CCGACTCTG TGTATAGGC 300  
CGTTGCGAG ACTCAGAGG GGCAGAGAC GCTAGTTTAT AGTAAACAG GAGCTCAGA 360  
GGTATATAG TCGAGAGAA CCGGAGGAG GTGAAACC TCCTGACTTA TCTCAGAGG 420  
AAGCGAGAT GACTTTTGA CTTTCTTAC ATGAGCTCTG TGTATAGAG TATATATCT 480  
GCTATATAT TACAGAGCA GTTCATCTG ATTTT 515

## 50 (2) INFORMATION FOR SEQ ID NO: 274:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2995 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

TCACACCAT AAGGATTTCA TGAAGAACT AGAGAGAGG GAGTGGAGG TTATCTCAC 60

AGTACCATG GAGATATCA TCTGCTCTT CTCACACCC GATTGATCA GATATCA 120  
AAGTGAAT ACCATATCC GAGCTGCTT CAGAGATCT CTCACATAGG CTAAACACA 180  
CGAGAGCTT CTTGAAAGG CTTTCTGAG ACTGTGCTT AATCTGAGC TCTGTAGAA 240  
ACTTGTGCA TGAATCATG GAGTGGAGC GAGCTCAT GAGGGAAT AGAGACCAAT 300  
CGCGAGAC ATTGACGAG TTAAAGCTT TATGCTGAG CATGACAT TTATGAGACA 360  
GATGATCTG AAAAGCTTG AGCTGAGAG GATCAGAGG AGTAAAGAA GAAAAACAT 420  
AGAGCTACT GAGGAGCTT TCAATGAGAA ATCCGACAG GAGGAGAGG AATCTTAG 480  
TGAAGCAC CTTCTTCCA TCCCATCTT TTCAAGCTT GAGCAGAAA ACCGAGAT 540  
GAGCAGCTT TCTTCCGCTT GGCAGAGCT GTGCTCTTA GATCTGAGC GGCAGAGCA 600  
AGTAAATAT GCGTGGATC GCTGAGAGG GTTGAAGCA TTTCGCACT TTGACTTTGA 660  
TGTCTGAGG AAAATGTATA TCGTTTGAAT GATTCAGAA AAGTCTGAG TGAATATTT 720  
CTTCCGAGC ATTGATAGG AGCAGATAG GAGATATCA GATCAGAGT TTATCTGATG 780  
CATTTTACA TCCAGTTTC CAGCAGCAA GTTGAAGAT ACTGCTGAG CTGACATTT 840  
CGACGAGAT GGGATATTT ACATTTATA TTATGAATTT GTGCTCTCT TTCAATCCA 900  
GAGAGATAG TATCAGACA GAGCAGTGC AGTAAATAT GATATAGAG TTACAGACA 960  
AGTGTCTAG TCGAATATG CAAAAATTT TCAATGAGG CAGTCTGAG AGATATATA 1020  
CGATTTCTT CTCGCAATC AATTTGAGA TTCTCAGAG TTGCGCTGAG TCCATATCT 1080  
GCGCAGCTT GATGTGCTG GTTGTGAGG GATGATATG CTTGATATA TTTTATATTA 1140  
AAAATATCT CTCGAGACA GAGTGAATA CTAACTTGA ACTTGAAGG AATTCATCC 1200  
TACAGAGAG AGATCTGAG GATATGACC CTTTCCCTC AGGAGATGA AGTTCAGAC 1260  
CATCTTCCG GGCAGCTCC CCACTCTTT CAGCTCTAG TCTATCTAG AGTAAACACA 1320  
GATATATCT TGTACAGCA CCGGAGCAG TGAACAGAG GTTATCCAT 1380  
CATCAGTAG GATTTTGA CAGACAGAC GACTTTTCA TTCTATCTG AGATCCCTTG 1440  
CTGTGATAC GAGCATTAG TTCTTCCCG GCTTCAAGG GTCCAGAAC TAAATGAGCA 1500  
GAGCTTAAA AGTGTGAGG TGGCTTGGG AGTGGGCTG GATGTGAGC GGGAGATTA 1560  
GCGAGAGC GCGAGAGAG TGAAGCTTCT GACTTTGAC TCTTGAAGC GATTTCTTG 1620  
TTTCGACAT TCGAAGACA GCGCTGAGG GGCAGAGGC AGCTCAGAG GAGGCTTAAA 1680  
GAGGCTTCC AAAATCTCA CATTCTTGA GAGAGACAG ACTGCTCTC CGAGAGCTC 1740  
AGTCCAGAG GATTAACAT GTCTAAGAC CCGGAGACA CTATCTACT TGAATCTTC 1800  
TGTATCATT GGTATATAT TTATCTTGA CCGAGAGAT TATATTTTA AAAATGTAAA 1860

AGAAATATTTG TGTATGAGG CTGGCTTATTT TTTTCTCTTT TGTATGATTA CTATTTTCTT 1920  
GTGATATTTT ATGTAGATTA ATTTGGCTC CTGGTAACC TGTATAGCT GGGGCCGAGA 1980  
5 AATGAATAT TTAGAAAAA CAATGAAA GGTCAAGATA CAATGTGTA TTAATAAAAA 2040  
AAAGCCATAT TATATGGTT TCTGGGGGT GAGGGTTGT AAACCTGTT TATCTTTTAT 2100  
GATATCTCT AATGACATCT TCTTATTA CTTGACTTC TATCTAGCA AGATATATTA 2160  
10 TATTAAAAA ATAGAAATCC TCGAGTGT TAAAGAACTCT TTTTGTGTA ATCAGGACA 2220  
CCTCAATTAG CAGAACTGA GGGGAGGGCT TTTTCCATG TTTAATGTT TGTGATTTT 2280  
15 AGCTAAGAG AGGGAACCTC ATCTAAGTAA CATTTGACA TGGATACAG AAAGGAGTT 2340  
CATTCGAATA CTGTCTTTGG ATATGTGTTT AGTACTGGT GTTTAAGGA CAATATGCTG 2400  
CTAGAATTC GGGTAAATG TAAATGTCA GAAACGTTCA GAACATTTGG GGTTTAAAC 2460  
20 TGAATTTTG CTCCTATCC ACCTAGACA CCAATTAATC TTGTGTTAC CAGGACCCAG 2520  
ACCTTGCCA AGGGATAGGC TGTGTGTGA CATTTGTAAT TCGATTTTG TTTTATCCAC 2580  
25 TTTTCTGCT ATTTATTTAA ATGTGCAATC AACTTCCAC AACTGAGGA ATGAATTCA 2640  
CGAGCTGTT CTGAATATGT GAGGTGAGA CAAGACGTG CTGCTCTCTT AATGAGTTC 2700  
30 ACCAGCAC TTGTAAACA GTCTGTGTTG CTTTGTGTTT TTTTGTGG TAATAAGTC 2760  
AAGTACCAA GTAGACATGA AAAGGGCTG TCTGGGGCT CTGTTTTTTA GTCTGTGTC 2820  
TTCAAGCTCG ACATGTTGC TGTGTATTA TCTCATTTGG TTTTATTTGA GGCAGAACT 2880  
35 GAAGCTCTAC CATGAACTG TTTAGAAACA AGACACACTT TGTATTTAA ATTGCTGCA 2940  
GTAAACAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAGGGGGGCC GTTAC 2995

(2) INFORMATION FOR SEQ ID NO: 275:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1990 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

GGAGCCCGCG CGSCTCCCGG CGATGTGTAG CAAGCCCGTG CTGCMKGTG TCTGCCGTCA 60  
ACCCGAGAG ATGAAGTGC TCTGGGGAT CGCCTTGCTG GCTTAGCTCG CCTCTGTTTG 120  
55 GGGCAACTTC GTTATATGA GGTCTATCCA GGAATATGCT GAATCTAAAA TTAGAAGCA 180  
GATTTAGAG ATGTTTGAAC CACTTAGAGA GAATATCAGA GATTTAGAA AAGCTTTTAC 240  
60 CCAGAAATAC CCACAGTAA AGTTTTTATC AGAAAGGAT CGGAAGAA TTTTGATTA 300

5 CAGAGGGCG AGKGTTCGTG GGTTCCTATC TTAATGTACA AACTCATGAT GGAGGGCAC 360  
GAGGTGACCG TGGTGGACAA TTTCTTCAGG GGCAGGAGA GAAAGGTGGA GCACGTGATC 420  
GGACATAGA ACTTCGAGTT GATTAACCAAC GAGCTGTGGG AGCCCTCTTA CATCGAGTTT 480  
GACAGATAT ACCATCTGGC ATCTCCAGCC TCCCTTCCAA ACTACATGTA TAAATCTATC 540  
10 AAGACTTAA AGACCAATAC GATTTGGACA TTAAACATGT TGGGGCTGSC AAAGGAGTC 600  
GGTGGCGTGC TGGTCTCTGGC CTCACATCG GAGCTGTATG GAGATCTCTGA AGTCCACCTT 660  
CAAAATGAGG ATTAATCTGGG CCAAGTGAAT CCAATAGAAC CTGGGGCTTG CTACATGAAA 720  
15 GGCAAAGCTG TTCCAGAGAC CATGTCTCTAT GCTTACATGA AGCAGGAGG CATTGAGTGT 780  
CGAGTGGCCA GAATCTTCAA CACCTTTGGG CCACGCCATCC ACATGAACGA TGGGGGAGTA 840  
GTCAAGACT TCATCTCTGCA GGGCTCTCAG GGGGAGCCAC TCACGATATA CGGATCCGGG 900  
20 TCTCAGACA GGGGTCTCCA GTAGCTCAGC GATCTAGTGA ATGGCTCTGT GGTCTCTATG 960  
AAGCAGAG TCAGCAGGCC GGTCAAGCTG GGGAAACCCAG AAGAACACAC AATCTAGAAA 1020  
25 TTTGCTCAT TAAATTAATA CCTTGTGTTT AGCGAAGTG AAATTCAGTT TCTCTCCGAA 1080  
GCCAGAGTG ACCACAGAA AAGAAACCA GACATCAAAA AAGCAAGCT GATCTCGGG 1140  
30 TGGAGGCCG TGGTCCCGCT GAGGAGGT TTAACCAAG CAATTCACTA CTTCCGTAAA 1200  
GAATCGAGT ACCAGGCCAAA TAAATCAGTAC ATCCCAAAAC CAAGCCCTGC CAGAATAAG 1260  
AAAGAGGGA CTCGCCACAG CTGAACTCTT CACTTTTAGG ACACAGACT ACCATGTATC 1320  
35 ACTGATGGG ATGTATTTTT GCTTTTTTTT TCTTGTGTTT TAAAGAAAGA CTTTAAACGG 1380  
TGTATGAG AACAAGCTGG AATTTCAATC TGAAGCTTCC TTTAATGAAA TGGATGTGCC 1440  
40 TAAAGCTGC CCTCAAAAA CTGCAAGTTT TGCCTTGCAC TTTTGAATC TCTCTTTTAA 1500  
TGTAAATAG CTTAGATCCA TCTCTCGTA TTTTCAGTT TTTTATCTT GCTGTGAGAG 1560  
45 CATATGTTGT GACTGTGTTT GACAGTTTAA TTTACTGTTT TCTTGTGAA GCTGAAAGG 1620  
AAGCTTAAG GGGACAAAAA ATCCGATTT TATTTATAAA AGTGGTACT TAAATATGA 1680  
GTCTTATAC TATCTATTA GAAATATCTT AGCAGTATTT TCAAGTCTG GTCCCGCGCC 1740  
50 ATTGATTTTA GGGCAGTAA AAGATTTCTG TGTGAGAGCT TTAATGTTCT CTTTAAATTC 1800  
AGAGTTTTTC CAGGTCTAC TTTTGAATG CAAACTTGAC TTTGAATAT TCTGTGTTGT 1860  
55 CATGATCAG GATATTTGAA ATCACTACTG TTTTGTCTG GATATCTGG GCGGGGGCAG 1920  
GTTGGGGGGC ACAAGTTAA CATATCTCTG GTTAACCATG GTTAATATAT CTATTTTAAAT 1980  
AAGATTTGA 1990



## (2) INFORMATION FOR SEQ ID NO: 276:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2436 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

10 AACCTCCCTT AGCTCCAG GGTAAACCG GTGAGNCCTT AAAACACAGA GAGACAGA 60  
15 TTTAAAGTCC GTTCATTGA AAATACAA CATTTCAT GTTTATCA AGGATCTCTT 120  
CGACATCCCT CCTCTTATA AGAGCAGCT AACATATCC TGAACCTTG TACAAAAGT 180  
TGAATTCGG CAAGACAG CAGTGAAGA TACACATCA GATTACACG CCAAGAAATC 240  
20 TTACACAGA CAAAAGAG ATCCAGGCA GATTATAC CCTCCAGTG GAAATATAG 300  
CAACATTTG GGCACCTTA ATTATGACA GAGAGAGCC TTGAGGGGAA GTAGAGTGG 360  
25 CCGAGGTGG GGCACAGAG GAATGCTAG TCGGGAGAG CTCTACTGA TAAAGATCA 420  
GCATTTCA GATTTCAT GACCTATA TACTTAAT CTACTTCA TTGATTCGC 480  
GGGATGTC CTTAACAG ACTGCTCT TCACTAAA ACTTAATGT CTTATACCT 540  
30 TTGTATGAT GACTACTT TGTAAAGAC CATTGTTGG TCGAGGTAA AACACAGTG 600  
AATTTTGG ATGCTGTC TCAATCTG ACTGTTTT GCAATTCAT TATTCAGCT 660  
35 TCAATTTGG AATCTTTAA AACTTTGAT AATTGTTGT TCAAGCTGT TCACTTTAA 720  
ATGATACAA ATTCATGTA GTTCTGTA TAAAGATCA CAGTTTGA AGGTTACGA 780  
TTTTCTCT CCTCTAGT TTTTACCA ATATTCAG AAGATTAAG GTCAATCTA 840  
40 AATTTGTT TTAATGTT AATTAACAG CTGGCAGTG GTCCAGCAT CTCTACTAGT 900  
GTCTAAAG CAATTAAT ACATAGCT CTTAAATAT AAGATACGA TTACATTTT 960  
45 AAGAGAAAT AATGCTTT GCAAGGCTA CTTAATCTT TTCAATAT TGTATACAA 1020  
ACTTTGAT ATGAAATCT ACTATTGA TGAATATG TATGTATAT ATACATCAT 1080  
ACATACAT AATATGAT GTATGTGT ATATATAT ATATGATG TGTAAAGCT 1140  
50 GACTACAA CATTAATC TTTTAAAT CAGAGACG GTAGTGCAC AGCGATATA 1200  
TCTTTGAG GCTGATCG TTAATACT GTTATTAG TTTTACTC CAGTACAG 1260  
GGATTCAG TTAGTCAAC TTAGCATT ATTGTATT AAACATGA AATAAGGTC 1320  
55 CATTTCAA GATTAATGG AATGCTGT GTTAATTA CACCAAAAT ACTGATTCG 1380  
ATGTACATC AGGTTCTAC AAGAAATAT GTATTAATT ACAATTTGA GATTATTA 1440

## (2) INFORMATION FOR SEQ ID NO: 277:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 782 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

45 GGCATCAT TCTCCACC TTCTGTCC CCAATATAG TTTATTTGT TGTGTGAGC 60  
50 TCACTTGG CCTTATTA AATTCATAG GAGCTGAG AAGCAATTC TACTGAGAG 120  
GTTAGAGC ACTTAGACA GAGCCGACA TCCCAATCT GAGATTTG GTGAGAGAG 180  
GACTTCTGG GGAATAGAC CAGACAAAT AACAGACT GACTAGAA GAGAGACTG 240  
55 TACAGATTT AATAGTCCA AAATGGTTA TATCCCTCC CCTTTAAT CAGATCTTG 300  
TGAATGGA AAACACGA AGAGGGGAT CAAGATGAC TGAATGCA TCTTTCCAG 360  
60 GAGGGGAA GTTGAAGTC AAACCGGCT GACAGGTGG TGAAGAGCC TGTTAGAGT 420

505

G

961

5 TGTGGCTGAT CCTCTCTGG TATTAGTTTT TCCCTGGG GAGGAGGC CTAGGAGAG 480  
GGGACTGCAG GTTCCCGCG GGAATTTC TCCCTCCCT GATGAGGCA GAGCGAGCT 540  
GGCTGCCAAG CCGCTCCCTC AGGAATGCC CTGCGCCAGG AATGCCACC ACACATACC 600  
TCTCTTTTT TTCTAGTCAA ACTCTGTTT ATTCTTGGC TTGCTTCCT CCTTCTCC 660  
CTCTCAAGCT TTACTCTGA TTCTCTTTC ATGGAATTG GATTCAGT TAACTACAA 720  
CAGTCCGCC AACACCAAT CTTCAGGAA AAAATACAA AGAATTTAA CAAAAAAA 780  
AA 782

(2) INFORMATION FOR SEQ ID NO: 278:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 961 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

30 GAGTTCCGCG TGAGACCG TGCTCTGGC CGCGCCTTC ACATGGCT CGCGAGCT 60  
GGACTACAC ATGAGATCC CGATCAGCC CTCTGGAGC CAGAGAAC GCGCCAGCC 120  
AGGTGGGAG GAGCGAGAA CTGCGAGCC TGTGTGATT CTTTGGCT GGGTGGCTG 180  
CAGGACAAG AACCTGCCA AGTACAGTC CATCTACCAC AAAAGGGCT GCATCTAAT 240  
CGGATACAA GCGCCGTGC ACATGGCTT CTCTCCGAG TCATGGGTA TCCCTTCACT 300  
TGTGTTTTG GCGCAGAGC TGTCTGAGCT GCTCTTTGAT TATGAGATTG AGAGAGCC 360  
CCTGCTCTTC CATGCTTCA GCACGGTGG CGTATGCTG TACCGCTAGC TCTCGAGCT 420  
CCTCGAGCC CTGCGCTTCT GCGCGCTGCG TGTGTGGGC ACATCTTTG ACAGCGCTC 480  
TGTGTAGAGC AACCTGTAG GGGCTCTGG GCGCTGTGCA GCGATCTGG AGCGCGGGC 540  
CGCATGCTG CGCGTGTGC TGTGTGTGC CTTTGGCTG GTGTGCTGC TGTTCAGCT 600  
CCTCTGCT CCTATCAGC CGCTCTTCA CACCCACTTC TATGACAGC TACAGGAGC 660  
GGGCTCTGCG TGGCGCGAG TCTACTCTA YTCAGGGCT GAGGAGTAG TCTTGGCAG 720  
AGACTAGAA CGCATGTGG AGCGAGCTT GCGCGCGCG GTCTGGCGC GTTCTGTGA 780  
55 TTTCGTGTA TCTGACAGC TCAGCCACT CCGTGACTAC CCGTGACTACT ACAGAGCT 840  
CTGTGTGAG TTATCGCA ACTGGTGGC CTGCTGAGC CATTCCTCA TCTCACTCT 900  
GCTCGAANA TAAATGGCTG ACATGCCC AAAAAAAAAA AAAAAAAAAA ACTCGAGGG 960

60

506

5

(2) INFORMATION FOR SEQ ID NO: 279:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1228 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

15 CGCGCTTGC AGTTGGTCT CCTGTGTAC GGCACGCC AGTAGGGGA TTGGCTTCC 60  
TCCAGTGGCA GCGCTATCAG ATTGGATAT GTCTTCATA TTGATTGGA TTACAGTGG 120  
TTTCACAGT GTCTACAGT TTTAGGANT ATATAGAAA ACTGTAAAC TGTATTTCT 180  
TGGATTGGT AATCCAGAA AAGACATTT GTTACAGTG CTAAAGATG ACAGCTTGG 240  
ACACATGTC CCACATTTAC ATCCACTTC CGAGAACTG ACCATGCTG GCATGACTT 300  
TACACTTTT GATCTGGTG GACATGTCA AGCTCGAGA GTGTGAAA ACTACCTTCC 360  
TCTATCAT GCGATGAT TTCTGTGGA TTGTGAGAC CACGAAGGC TGTAGAGTC 420  
AAAAGAGAA CTGTCTAC TATGACAGA TGAACCTTT GCTATGTGC CTATCTGAT 480  
TCTTGGGAT AAGATCGACA GACCTGAGC CATCAGTGAA GAGAGTTGC GAGAGATGT 540  
TGTATTAT GTTCAGCAA CAGGAAGGG GAGTATATCT CTGAAGAGAC TGAATGCCG 600  
ACCTTAGAA GTTTTCATGT GTATGTGCT CAAGAGCAA GTTACGGAG AAGCTTCCG 660  
CTGATGGCA CAGTACATG ATTAACAAA ACTCAGATTG GTTCCAGTC TCAAGCTTCA 720  
GGCTTACTCA GAGATTGAT TCTCAACAT GCATGACTTG AATCAATAG ACTTTTCTG 780  
GTTTAAAC AGATGTTTTT TAGATTATTA ATTTAAATC ACTTATTTT GAATGAGAT 840  
TGAANACTGA TTCAAGTAG TTTCAGTATC ACATGTTAG CTTTCTAAT CCAATAAGT 900  
ACTTGTGTTT TACAGTTAT ATCTGACAT CAGCCGAGCG CCATTTGTAA AGAGCACTT 960  
TCCAGGAGTA CATTGAGC ACTTTTAAAC AAGTGAAGC TATTAACCAT ATTAAAGAC 1020  
TCATCATGTT AATTTTTTA TGTACTTTC TCGAAGTCT TTTTAATTT TAGATTATAT 1080  
GTCCACCTAT CTTAGTGTGA CAGTAAATA TTAGCTTATT CAATGATTC ATGATGCTT 1140  
ACAGTTTCA ATACTTTTT TTCTATGCA AAGTCAATC AATAAACA ACTCTAATGT 1200  
TTGCGAANA AAAAAAAAAA AANTGCA 1228

60

## (2) INFORMATION FOR SEQ ID NO: 280:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1327 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

10 TCTGGGCTCT CCGGACAGCT GAGCAGCTTG ATGAAAGCCA GGGTCTGAT GGGGACCTTG  
GGGCGGTGCA GAGATCTGTG GGGGCGCTTC GCGAGGGGGS CCGAGACCGG TTAAGAGTGA  
120 TTCTGATTT TTAATGAGCC TTGAGCAGCT TTGCATATTA TCGAAAGCCA TGCCCTTCTT  
CTTACATAT TTGCGATTAAT AGCAGATCA TCAATGAGGA GTGTGGAAA GAGCTGACAG  
140 GGTCTCTCA CCAATTTAC CCAATTGAGA TCGACCCACA CCGGACCTTC AAGGAGAAC  
TACCTATAT GTTGAAATGG TCGACGAAAG CCGACATAT CTATGTGAG CAGAGATATC  
160 AAGATTTCA GATAGCCGAG GTGTGTAGAG AGTCGATTC AATGTCTAG GAGGATATTA  
AAGCTCTT CAGACAGCTC TACCATACAA ACATTCCTCT TTTCATCTTT TCTGCGGGCA  
180 TTGGTATAT CTTGAGGAAA ATTATCCGAC AGATGAAGT GTTCCACCC CAAATCCACA  
TCTGTCTTAA CTACATGAT TTTAATGAAAG ATGGTTTCT CAGAGATTT AAGGCGCAGC  
200 TGTATACAC ATACAGAGAG AACAGCTTG TGTGTAGAA CTGTGGTTAC TTCAGACAC  
TTGAGGCGAA AACCATATC ATCTCTCTGG GAGACTTAT CCGGACCTTC ACATGGCGG  
220 ATGGGGTTTC TGGTGTGAG AACATTTCA AATTTGGCTT CTGTATATAC AAGGTGAGAG  
AAGCGCGGGA AGCGATTAAT GAGCTCTAT GACATCTGC TCGAAGAGGA CAGAGCTTG  
240 GATGTGTCA ACGGCTTACT GCGACAGATC CTGTCCGAG GGGGTCCAG TCGAATGCA  
AGGCGCTTGA AAGCGAGAGC TCCAAAGACC CTCGAGAGCC GTGTGAGGA GGGGCGCTTC  
260 CCGAGATCT GCTCCCCCTT GAAACAGAG CAGAGCGAG GTGGCGAGCA GTGCTGTGTT  
CGTTCCGAGC CCGTCCGCTC TCTTTTCTCT GAGCAGCTTC ATCCAGAGG GCTTGAGGA  
280 AACCCGCAAT GTGGAGAGGC ACGAGCAGTG TTCTGTGTGA ACTTGTGAC AACAGATTC  
AGTGTCTAG GATTTGTCTA CTGAGAGGAT TTTCCTGAAA ATTATTAAAC ATGGAGATTT  
300 GAAAGAAAAT TAAATGTGTA AACAGATCA AATTTTAAA ATGAAAAAAA AACTGCTCTG  
ATTGAGAGGA TTGTGGCTGG GTTGAAGCT GAGCTCTTG GCTGGGAGGC AATGGAGATG  
320 CTCTGAG 1327

60 (2) INFORMATION FOR SEQ ID NO: 281:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 799 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

10 TCAACCTGCG TACAGCTGAG ACTGAGAGC ACTGCGCCCT CAGAGCTGAC TGTGAGAGG  
TGATATTCAC AACCTGAGAG ACCCTGAGAG CAGACCTGAG GGTGTCTCCC GTGACGTGT  
120 GAGTTGAGT GAGCTTAAT TCGACAGCTC CTGTCTTTT GAGAGAGCT GAGAGAGCA  
TAACTCTGC CTGCTGAGCC CAGCTGGGCG CTGAGTGTCT TGTAAATGCA TTACTTTGCG  
140 ATGTGGGTT ATTCCAGCCA AAGCATTTTC AAGTGTCTGT TAATGATTTG TACATATTTA  
TAAATATTA TCGAAGATTT GTTCCATTA TCGAGCTCT TTGCTCTGGG TACAGCCACA  
160 CCGCTTAC CCGACCTTGG ACTTGAAGAC CTACGTGAG GAGGTTTCC AGGTGTCTCT  
AAGAGAGAT TCGTGAATCT AGCTGTGAC GAGCATTTT TCGACAGAGT CAGAGAGCA  
180 TTGCGGCT GATGGGTTAG AAGTTGGTT TGCTCTTGT TTGAGCCCAA TATGTAGAAA  
ACATTTGAA CAGTTGAC CTGTATAGAG GTATGATTT TCGAAGCA CCAATTCATT  
200 TGTGATATT TATGTGTG TGCTTTATTA ATCATATTA CAGATTAAT ACCTTTCTT  
CAATTTGCT TCGAGAGAC ATACTTAG TGTTTTTGT TTGTTTTTG TTTTCTTCT  
220 TTGTTTTCT CTGTATAGAG AAAAAATTA ATATCATAT TTTAATATC AAAAAATGCA  
CTAAAAAT CAGGAGGAG 799

## (2) INFORMATION FOR SEQ ID NO: 282:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2196 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

50 AAGAGCTTA AATTCATGTA GCTTGAACAT GAGCAGAGAC CTACTGTGCG CAGCGAATG  
GCTGAGCCCT TCGGTACTCT CCGAGATGGA TGGGTCTCTT ACTAGACCA GCTGTCTTTT  
120 CTGAGCCACA TGGGCTTGC TTCTCTTAT ATGAGCTGCC TGGGCTTTGA CTGCATACCC  
180 AAGGTTAGC CCGACATCA GAGAGTAAAT GGTTCGATCC TCAATATTTT GATGAGAGCA  
TCACTATTA CTGAATTAAT GCGAATGTA GCTTTTACT GCTTACGTG AAAAAATGAT  
300

60

5 TGGTTGCGA CAGGTCTGAT CTCAGGNTG GCACAGCTTT CTTGTTGAT CTTGTGTGTG 360  
ATCTCTGAT TCATGCTGG AAGCCCCGTG GACTGTGCG TTTCCTCTTT TGAAGATATC 420  
CGATCAAGGT TCATTCAAGG AGATGCAATT ACACCTACCA AGATACCTGA ANTACAACT 480  
GAATATACA TGTCTAATGG GTCTAATTTCT GCTAATATG TCCGGAGAC AGTCTCTGAA 540  
TCGTGCGCCA TAACTCTGTT CAGTCTGCTG TTTCGAGGG TCATTGCTGC TAGAATGCT 600  
CTTTGCTGCT TGCATTTAAC TGTACACAG TTGCTGCAAG AAAATGTAT TGAATCTGAA 660  
AGAGGCATTA TAAATGCTGT AGGAACCTCC ATGAACATATC TTCTTGATCT TCTGCATTTT 720  
15 ATCATGTGCA TCTTGCTGCC AAATCTCGA GCTTTTGCT TCTGTGTAT GATTTCAATC 760  
TCTTTTGTGG CAATGGGCCA CATTAATGAT TTCCGATTTG CCCAAATATC TCTGGGAAC 840  
AAGCTCTTTG CTTGCGTCCC TGAATGAAA GAATTTAGGA AGGAATACA AGCAATACA 900  
TCGTGCTTTT GAGACAGTTT AACTGTGCTT ATCTGTTTAC TAGATTATAT AGACACATG 960  
TCTTAATTTT GTACTGCAGA ATCCATPAA ATGCTGTGCT GTTTTGCTT GTTTTACCA 1020  
25 CAGCTGTGCC TTGAGACTA AAGCTGTGTT AGGAACCTA AGTCAGAGA AATTAAGTA 1080  
TTAATTTCCC TTAATGTTAG GCATGGAAA AAAATTTGAA AAAAAAAT CAGTTTAAT 1140  
ACGGAGACTA TAAATGTAAC ACTGAATCC CTAATTTCTC ATGATGATG ACAATCTTAC 1200  
GTAAAGAGCT GGTTAGTAC GTGAATTCAG TTATCAATTTG ACAGATCTTT ATCTGTACTA 1260  
GAATTCAGAT ATCTCAGTTT TGTGCMAAC TCACTCTGTT TCAAGACTAG CTAATTTAT 1320  
35 TTTTGTGATC TTAGTTATTT TTAANAACA ATCTCTCAG TATGAGACT AAATTTTAT 1380  
AACTAATPAT ATCTTATTTG ATCTAATGA TCTTAAGGTA TTATCATGTA TGTGAAAA 1440  
CAAAACACTT AACTAGATTT CTTAATAG GTTTATGCTT TAGCTTAAG AGCACTTTG 1500  
TATTTTATTT ATCAGNTGGG GCAACATATT GTATGAGCA TATGTAGCAT TTCAAGCAT 1560  
GTTATCATG TAACTGTGAG GTAGAGCAA ACCTGTAAAG TAGATTATTC ACACATGAC 1620  
45 TGCATACGA CTTCAAAATAT GTCAATGTT TGTGTATAGA ACCTAGAGC CAAGAAGCAC 1680  
ACGAGAGGCG AGAATCCCA ATTTAATCTA TGTATCATC ATTATGATC TGTGTGTATG 1740  
AACAAGGG TGTAAAGCTT CAGCTGTGCA AGTTACATGT AGAAGGCCA CACTGTGAA 1800  
50 GGTTTTGTTT TACAATAC TTGAATTAAC ACACCTAGGT AGAATATTTT TATTTTACT 1860  
GTTTATPAC CAGATGTTAT TTCTAGATG TTCTACGCA AGAATATCA TAAAGATATC 1920  
55 CTTTTCAAAT GCCTTTGAGA AGATAGAG AAAAAGATT TGTATATTT TTAATAAAT 1980  
GTTTAAAG TCAGTTTCCA ACATGTGCTT ACCAGATGG TACTTTGCTT TAACGTTTAA 2040  
60 TATGCACTTT CATGAGACT GAAATAGCTT GCTATAGCA CTTTCTTAT CTTGAGATT 2100

5 TAACTCTTGG CTTCACTTTT CTACAGTATG ACATAATGAT TTCTATATCTT GTAAAAATCTT 2160  
TGTAAAAAT TTCTATATAA AATATTGAA ACTTAA 2196

(2) INFORMATION FOR SEQ ID NO: 283:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1185 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

60 GCAGTTAAGG CTTCTGNTAA GGAAAGAGAG TCTGAAACGA GCACACACAT CTGAGACTCC  
120 AGAGTGGGG GATCCAGCAT CAGATTCAT CTTGAATTTT TCTTAATAA CTTTGTACTC  
180 ATAAATGATC TCAACAAAGA TCTGTATTTT ATCTGTGCTT CCACTTTCCC TCTGGGTCAA  
240 GTAGATGTTA AGCTGGACCT TGGCAGGCTT CTTAAACATGA AGAATCTAG CTAGACAGAC  
300 AGACTCCCC ATTTATGGGA ACGAATATC AATTAATCTT CTAATTATTA AACATTTTIT  
360 TAAATGCTCT TGGTTTAAA AATCTAATATG TCTGGGGTGT GATCACTCAG GAGCACTTAA  
420 CTATCACTCT TCGCATGCTT TGGTCACTGG GAGATCTCTT GGGGCTGGG AGGTCTCTCT  
480 GTCCAGGCT AAAGGAAAAG CTTCAAGGG GTAAAGGCCA CAGAACCTTC GGCAGAAAG  
540 CCGGTCTAGG GAGATGTAAT GGTACAGAGA GGAAGGAGAG GAAGGGGGGT GGAACAGAGG  
600 TAGAGGCCA GGAAGGGATG CCGCACTGGA GACCGATGGG GACACTCTAA TTGTGCAAGA  
660 GGGAGGATCT TCTTCTTTGA ATGCTGAACA CAGCTAGTCT GAACCTTCTT TGGAAATGTC  
720 AGCTGTTTGC CCAATGCTAG GGCACACTCT CCGTGCAAAG CAGCAATATG GCGTCTTATC  
780 AGAAGGAAA AGTATCCATC AGTGTGACAA GAGGTCACTT TCGAATCTGC ATGAACTCCT  
840 TCGCGAGCCA CAAGAGTCC TGGTAGAGT GAGGATGCC TAGTCTTAGG CCGTGTCCGTT  
900 TATAGAGTA GCGATACAA ACCTCTGCTA GTCTCTGAAA TACAAACAG ATTGTAGATC  
960 CACTGTCCA TATGAGCTG TTGAGTTTIT TCCAGCTTAA GTTCATGACC CAGACATGAA  
1020 GGGAGATCT GAGGGCAAG TACACAGCTG TCAGGATGAT GGTCCCTTTG AACTATATGA  
1080 ATAGAGGCTT GACCAAGCCA GCGTGGAGGA CGAAGGTGTT GAAGAATATG AGGAATATGA  
1140 TGAATATCTT GAGAGGACT GCAATATCTT GATGCACTG AGGAGAGGT TTCTAGTTCC  
1185 TTTGATGAG AGCTGTTTCC CTTCTCTTAA GCGAGGCC TCCTA

## (2) INFORMATION FOR SEQ ID NO: 284:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1634 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

10 AGCGAAGAGG GAGGCGAGCG GAGGCGAGCG CTGAGCGCTG AAGCTTTAGC 60  
AGCGAGCTCG TCGAGTCCCG TGTGATCGAG AGCAAGCCAG GCTGCTGCTG GCTGCTGAGC 120  
15 AAGCTTGGTT TGTACAAAG CAGGCTGGGA GTCTATTTT GTACATGAGA TACATCAGAC 180  
TTACCTGTG GCGAGTATG TGAAGTAGT CTGAGTTGTT TACAGTAGG CTTTCCCTGC 240  
CGACAGAAA TTGTGTACAT AGCTCTGAGA TGTATCAGCC CTTTCCCGA GCTCCAGAC 300  
20 AAGAGCTGCT TGTAGGCTCG TGTATATGT CATATTAGC GTTTTATAT ATGAGCTTTG 360  
ATTTCGTGCT TTGTATTTT AGCAGAGTGT ATGAGCTTTC ATTATATATC ATCTGTGTC 420  
25 ATACAGATAC GAGTATATGT GTGTGCGTAT GAGTATATCT CTGATCTGTA GTTTCAGAGA 480  
GTTCAGCTGA AGCAGATGGA GTTCTGAGAC GCGAGAGAGA CCGTGCATCC CTGCTATATG 540  
TGTTCGCGAC AAGATATAGT GAGTCTTCCCT TATTTATAT TTGATTTTCA AAGACTTGAG 600  
30 CAAAGCTGAT AGTGTGTGCT GTTCTTTTGG CAGCTAAGTG AAGATCTTGG GATGACTTGC 660  
TGTGTCTCTC AAGCTGACT TTGGGGGCGAT CTGTGAGATA TTAGCCCCCT TTTTCTGTGG 720  
35 TGTATCTGCT TGTGTGCTCG TGTGTGTGCT TGTATCTGAC TCTTGCATGG CTTCAGATGTC 780  
TGTATTTTGG CATTTGGGGA TTAGGTGCTG AAGCCAGAC ATTGCAGATT TGTTTGAGGC 840  
40 CTGTGCGCA ATGATATATC ACTGCTTTTG ACTGTGTATG TCTGTGCTGT TGTCTCTTTT 900  
CTTGTCTTTC TCTTGAGAGA GGAAGAGACT CTGTGAGGC CAGAGCTGAG TGAAGTAGAC 960  
TGCAGCTGAG TCGATGACTT CTTTAGAGAG AGAGAGAGAT ATGTCAATTT ACTATAGTTC 1020  
45 TTAACAGACA TTATGAGAG CAGAGCTGCT TCGATATCCA GAAAGTGAAG CAGATATAGG 1080  
TTATATCTG CTCAAGATTA TGTATCTGCT AAGGGGTAAA TCAAGGCTAT ACATATCTG 1140  
TTTACGCGGC AAGATATAGC TGTATATTC TACAGCTTTT TTTTATAG CATTTACTTT 1200  
50 TTGAGCTTTT CCTCTGAAAG TCGAAGAGGC GTTACACTTT CCCAATGTGA GATGAGATC 1260  
TGCAGATGAC GAGCAGCTGT ATGAGTCTGC TTTCCAGAG AGAGAGAGCT TTTTGAAGAC 1320  
15 AATATGCTT TATTTGTATT GCGCAGCCCT GAGCTTTTGG GTTGAAGAT TGCAGATGAG 1380  
AATGATTTT AAGAGAGAT GCTTGAAGC ATGGGTATAC AGAAGAGCT ACCTAATCTC 1440  
60 AATATCTGC AAGCAGAGA GCGACAGAG ATTACTATGC AGCAGAGAA TATTTATAT 1500

TGAGTCTGCT TGTATCCAAA ACTGAGAGAC CAGTGTCTTT GAAATAGATC GAGCTCAGAG 1560

CTGGGGCGCG TGCCTGAGAC CTGTATATCC AGCATTTTGG GAGAGCGTAA GCGCGGGCGCG 1620

5 KTTACCGGG GGTCT 1634

## (2) INFORMATION FOR SEQ ID NO: 285:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1795 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

10 TTTCCCCGAG GTTGCCTTCC TTGCATCTCT TTTCTGTGTA TCAAGCTTGG ATTCAGAGAA 60  
20 CAGCCCCCTC TTTGTGACAC TCAATATGTA GCTACATGAG GAGAGAGAGC TCGAGCGAGAG 120  
15 CTGAGCTTGA TGTATATATGT TACCTCCGAA GAGCGGAAAA ACTGTGTGTG GAGCAGAGAT 180  
GAGCTCTGAG CCGAGATCTT GAGAGAGCTT GTGTGCTTGG CAGCTGAGAG CCGTCCCATG 240  
25 TTAGAGAGAC AGCTGATGGA TCCCGGGGGA CTTGGGAGCA TCAAGAGAT GTTTCGCGCG 300  
30 CCGTGTGACA TTATAGAGAT GCTGATATGC CTGTATCTCT GCGATATCCC GCGGCGAGGC 360  
AGGCTTTTGG ACTGCGAGAT TGCCTCTTTC TGCCTGGGCGC TGCCTGAGCA GCGCGGGGCGC 420  
35 TGAATCTGTA TGCCTGCTCT GAGTATATAT CTATCTGAGC TGTATCTGAG GAGACTGAG 480  
GAGAGCTTGG GAGATGAGC CCGTTTCTTC TATGAGAGAG ATGCTGTAGA GGTGATTTGCT 540  
40 GTTCTGTGGA AAGCCAGAG CTTCAGAGCG CAGCCCTTGA AAGCTCTGAG CAGCAAGAGG 600  
45 CCGATGTGTA TGTCTGAGAG TGGGAGCTTA GTTATGTGTC CCAATGTGTA AAGCAATCTG 660  
GAGAGCTTGG CTGTGCTGAG TGAAGGCTTG GTTGCAGATG TCGAGGCGCG AAGTGAAGAG 720  
TGCAGCTGCT GATTCGAGCT CTGAGAGAG CTGTGAGAG AGCAGAGAGC ATTGAGACTTC 780  
45 TTAGAGAGTA TGTGAGTATG ATGAGCTTGA CCGTCTTGG AAGATGATCC TCGATGAGAG 840  
GCTGTGAGC TGAAGATGCT GATTCATGTC CAGCAAGAG CAGCGCGAGC TTTCTGTGCT 900  
50 ATGCTCAGAG ATTTGGGAGAG GAGCAGTGTG GCGCATGTAG TGTCTGTGAG CTTCAGATCC 960  
CAGAGAGAGA GTTGGAGAGC GCTTCAGAGAA GAAACTGAG CAGAGAGATC CAGTCACTTA 1020  
TTTACGCTTG GCTGTGAGCT CCGTGCATAT TCCGAGCTCT TTTTATATGT TCTGTGAGAA 1080  
15 AAGAGAGAG GAGTGTGTC CTGAGAGAG CTGTGCTTCC TTCTCTGAGC CAGAGACTGT 1140  
GCTTGAATTT AGCATGCGCTT TCACTGAGAT GTGAGGCTTT TTAGATGAGAC CAGAGGAAAA 1200  
60 TGTGGCGCTT CTGAGTGA CAACAGAGAC TGAAGAGTGG AAGGGGCTTA TATGTGTATG 1260

5 AATGACCAC ATTGAAGAG CACAATGCC TCTGTGTG ATGCCACTTC CCAAGGTGA 1320  
GACAGTGGAA AGAACCCGAG GACAGGAAG GATTGGGTAG GTGAAGGGGT CAGGGGACTG 1380  
GTAGTCACCC AATCTTGAG AGGTGCAAA AGCACTGGGG GCTACCCGTT AGCTGCATCT 1440  
GCCCCGCTG TTTCGCCGTT CATGTACAA ACTGCCACTA CTATGTACCT GCAATGGGGT 1500  
TCCAGAGATG GGGGAGACTC AGCTTCTACT CCCAGGAGC TCCAGGGGCC CAGGAGAGAG 1560  
AATCTGCTT CTTTCAGTC TGGTCTACAC CCATTTCTG GTAGCCCTTC TCTTCTGCT 1620  
AATTCGCTT GTTTTTCGAG ACTCAGCTCA AATAGTGCCC CTCCTTTAGC CCATCCCTGG 1680  
CCCCAGGCT GAGGTGATCT TTCCCTCTTC TGAATATTTA GAGGATTTAC TGTCTGTCA 1740  
GTTCGTTGG CAGGCACACA CAGTGGGATA AATTCATTTG TTTTGAATCT TGTATT 1795

20

## (2) INFORMATION FOR SEQ ID NO: 286:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 858 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

30 TCTCTTTGG GTCTGTGGTG TACTCTGGG CGCTCTCTC CGCTGTGGG GTTTGGCCA 60  
GCTCTGGAG GAGATCTGG CTCAGTTTC GGAGCGGATG AATGCCCTGT TGTGCAAT 120  
TCTGAGGTG TTCCCGCTGA AGTATTTTGG CTACAGCCA GATCCCTGA ACTACCAAT 180  
AGCTGTGGC TTTCGTGAC TCTGTGCTGG GTTCTGTCTG GTACAGGCC CACGATGCT 240  
GCAGAGATC ACTAATCTCT TCTTGATCTT GCTCATGATG GGGGCTATCT TCACCTTGG 300  
AGCTCTGAA GAGTCACTAA GCACCTGTAT CCCAGCCAT TGTCTGTGG GTTTCTGCT 360  
GCTCTGTAAT GTGGGCCAGC TCTTAGCCCA GACTAGAGAG GTGTGAGAC CCAGTAGGA 420  
GAGACTCTA AGTACATTTA AGGAATCTTG GAAGTAGAGC ATCTCTGTCT CTTTATGCA 480  
TCCAGCTGT ACAGCAGAA CATGTAGAA CACAGATCT ATCATCTGT TACAGTATA 540  
ATATCCAGG TCAGCCAGTG TTGAAGAGA CATTTGTCT ACTTGGACT GCTTCTCTT 600  
TTTAGCTTTA CTACTCTTTT GTGAGAGTA CATGTATGC ATATTAATCT TCTCATGTC 660  
ATATGAATAT ACMAATTAG CAGAAAGAA ATTTAATCA ACCMAATTC TGATGCCCA 720  
AATAACCACT TTATATGCT TGGTATAGT ATACTCTGA ACTTTTCTT GTGCCTTAA 780  
ACAGATATAT ATTTTTTTTT ATGAATAATA AAACATATA TCCATTTTAA TTCTCTCTT 840

60

TTAAACCTT ATAACTA

858

5

## (2) INFORMATION FOR SEQ ID NO: 287:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 915 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

15 GAAATGGCA CGAGCGGC CATGGGCTC CTGCTTTGG TGTGCTGT ACTGCTGGC 60  
GGCTCTTGG CGCTGTGGG GTTGGCAAG CTCTGGAGG AGATCTGGC TCCAGTTTGG 120  
GAGCGATGA ATGCCCTGTT GGTGAGTTT GCTGAGGTGT TCCGCTGAA GGTATTTGGC 180  
TACGAGCAG ATCCCTGAA CTACCAATA GCTGTGGCT TTCTGAACT GCTGCTGGG 240  
TTCTGCTGG TCATGGGCC ACCGATGCT GAGAGATCA GTAACTTGT CTTGATTTG 300  
CTCATGATG GGGCTATCTT CACCTTGGCA GCTCTGAAG AGTCACTAG CACCTGTATC 360  
CGAGCATTG TCTGCTGGG GTTCTGTCTG CTGCTGATG TCGGCCAGCT CTTAGCCAG 420  
ACTAGAAGG TGTGAGACC CACTAGAAG AAGACTCTAA GTACATTCAA GGAATCTGG 480  
AAGTAGACA TCTCTGTCT TTTATGCCAT GCAGCTGTCA CAGCAGAAC ATGTTAGAC 540  
ACAGAGCTA TCACTCTTT ACCAGTATA TATCCAGGT CAGCCAGCT TGAAGAGAC 600  
ATTTGTCTA CTTGGACTG CTTTCTCTT TTAGCTTTAC TACTCTTTG TGAGGAGTAC 660  
ATGTTATGA TATTAACAT CTTCAATGCA TATGAAATA CAAATTAAG AGAAAAGAA 720  
TTTAAATCA CCAAAATCT GATGCCCA ATACCACTT TTAATGCTT GGTGTAGTA 780  
TACCTCTGA CTTTCTTCTG TCCCTTAAA CAGATATATA TTTTITTTTA ATGAAATTA 840  
AACCATAT CTTATTTAT TTCTCTCTT TAAACCTTA TAACTATA HAAAAAAAAA 900  
AAAAAAAA CTGA 915

50

## (2) INFORMATION FOR SEQ ID NO: 288:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1517 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

60

CCCTGTGCGA ACTAGTGGGT CCCCCGGACT GCGAGGATTC GGGCGATGAT TCTGATCTG 60  
AAGATCTT GAGTTCCTCT GAGAGATCA AAGGCTCCGG GAGCGAGCC CCGAGCCCCA 120  
AAGCGAGCC TCGAGAGAC AGGAGAGACC CTGAGGTGAC CAGGGGTGAC CCTCGAGAGA 180  
CGAGCTTAG CCGTCAGAG ACCAGCGAGA GCGCTGAC GCTGAGAGCG ACCCGAGTGA 240  
CGAGCTTACT TGAAGCTTTC CCGAGAGAC AGTGTGACA GAGAGAGTG GCGAGAGTG 300  
GAGGCTGAG TCGTTCGAGA CAGGAGACA CCGAGGCAAT CTCTAGAGAG CTGAGCCGAC 360  
CTCGACCTTC ACCGTGTACT CAGAGCGACA GAGCGAAGG TTCTGACTCA AACTGAGTGC 420  
CAGAGATGCG GCGTGTTC AATGAGAGAA CTCTTTCGAG GGGGCGCGCA AGCTGTGCA 480  
AGTCAGAGAG TCGAGAGAGC TGTACTGAC CCGACTGTG GCGATCCGTA CCGGATGCG 540  
TTTCTGGTGT CAGCGAGACA AATCAGGTT CTGTGTGTTA CCGAGCTCG GAGAGAGCTT 600  
TGAATCGGCT CTGAGATCA GCGGAGAGCA TGTGCTGTGC AGAGAGCTT TGTCTGAGG 660  
TGGCTGTGCG GCTGTGAGAT GCGCTGAGAT TGTCTGAGA GAGTGAATAT GTTACGAGAA 720  
ATGTGAGAG TGAAGATATC TTGTGTGATC CAGAGAGCA GATGTAGGTG ACTTTGCGAG 780  
GCTAGGCTT GCGATTCGCC TATTCGCCAA GTGCGAAGCA CCGGCGCTTC GTGCGAAGCA 840  
GAGAGAGCT TCGAGAGGGG GAGCTTGAAT TTCAATAGCA TCGAGCTGCA CAGGAGATGC 900  
GGGCGCTGCC GCGGAGTGA CCGTCAGAC CTGGATTAT GCGATCGAA GTGTGTCTAG 960  
GCGTTTCTGC CATTGAGAAA TTGCTCTTCA AAGATGAGAA CATTGAGAG CAGAAAGCGA 1020  
AGTTCCTGTC GATTCATTTT TAAATGAGC TTAGACTTGT CATTGAGAA AGAGGCTTGC 1080  
GTGAGCTGCA GAGATTGCA GTTTTCGAC TCGACCGAGT TCGTCCGTGT GATTGAGCCA 1140  
GATATCTCG TTATGAGACC GCGTCCGGGG GAGCAGCTCC CTGCGTTTGA GTGAGCGACA 1200  
GAGAGCTAC TTGAGCGGCC CCGTGGCCCC GAGATTCAC TGAATGTGC GATGTGACA 1260  
GTGAGCCCTC GTGAGGACA GCGATGACTC TCGAGTGA GCGTCTTGC TTGAGACTTA 1320  
CGAGTTAAAA TGTCTCAAA ATGTGGAAT ACCGCGCAAT GCGCGTGC ACAGCAGAGC 1380  
TCCGTGACA GTGTGCTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GAGAGAGTG 1440  
GAGAGTGC CTGTGGGTGT GGGGTGCG CCGAGAGCT TGAATGATA AATCTTTAT 1500  
TTCCAAAAA AAAAAA 1517

55

(2) INFORMATION FOR SEQ ID NO: 289:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3865 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

5 TCGAGGGGG GAGCTTCTTT GAGCAGTGG CCGAGCGTC GCGCTGACA CTTCATCTC 60  
TGAAGCTTCT CTCTCTTCAAT TTGGTGCAT GTCTTTCTG GAGTGCCTT TCGAGAGAG 120  
TGGTTCATCT GGGGCGAGCT AACCTGAGT GAGAGAGTG GAGAGCGACA GATTCATTTT 180  
ACTGAGTCT TCTGTAGTCA ATGAGGGGCA CCGAGTCTT CTAGGGGAGG CTGGGTGGTG 240  
GTCCCTTAGG TATGAGCTTC TGTACTGTGA CTCTCGGAAA ATGTGAAGCT TTCTATTTTC 300  
GTCCTGTGCG TACGAGCTGC TCCCGCATTC GCGCGTGTG GTCCAGAGCA 360  
AGCTGTGCG ACCGTGTAG GCGAGTGC TTCCCGATTC GCGCGTGTG GTCCAGAGCA 420  
GCGTGCCTGC CCGCGAGGCG CAGCGCTCT TTCTGATCC TCTTTCTTGA AGAGTGAATT 480  
GCGCTGAGT CTGAGAGAG ACCTTCCTT TACCTTCAC AGCAGAGAGAG GAGGTGACA 540  
TGAAGTCCCC GCGCGTGC CAGAGCTGG AACAGAGGG ATGTGTGAG AGCAGAGTTC 600  
CTGTGGCTCT CTGAGAGTG TTTCCTACTA GTGACTGTC TCTTCTCTT GTAGCTATTC 660  
AATCAATAT CTTCCTTGC CTGTGGGAG TCGAGAGTG CTGTGGTG TACGTGCAC 720  
CTGCGAGTG AGTGGGGAA AGAGATTAAT CAGTGAAGC TGTGTCTGC AGAGCTCTG 780  
ATCTAGCCCA CCGCTGAG TCGAGAGTG GTTCAGCT GATGAGAGC AGCGCTTGC 840  
AGCAGCTTC GAGTGTGTC GAGTGTGAG AGAGCTGAC TTCTCTTTC CTCTCCCTCC 900  
TTCCAGTTTA CTGAGACTT ATCTGTTAG GATCTGTGA GCTGTGTTC CTGTGGGTG 960  
GAGAGAGAG CAGAGAGAAA GGGAGGTCT AGAGAGAGCA GCGCTTTTTC GTCTCTGCG 1020  
GTAAATGAG TTGAGCTGAG GTAAATGAG AGACGAGAG CCGTGAATT TTAATTCGA 1080  
TAAATGTGA GAGTATTA TATGATTA TATATTTCT TAAATTTTC AGCTTGAAT 1140  
ATGTCAAAA ATCCATTCGC TGTCCCTGAG AGCTGAGTG AGACATGA AGAAAGCTT 1200  
GTTCATTTTA AAGATGTGA TTAATGAT GAGACTTGC TGTGGCTGAT GCTTGTAT 1260  
GTTCGGGGGA CAGGAGAGTG GTTCGGGCC ACATTAGAA GGGAAATGT TTTCCTCT 1320  
GAGAGATTC GAGCGAGTGA TGGGCTCTT CTGCTGATTA CTGCGAAGG GACTTTAG 1380  
GTGTCTTTC CATTCTTAC CCGCAGCC CCGATTACGG TAAAGGAGAG CCGAGAGTG 1440  
GTGAGGGGCG AGAGCTGCG TCGATGTG TTCTCTCTG AGATGTGCA GCTGAGATCC 1500  
AGAGCTTC TTGAGAGAT TTATTTGAT TTATGAGCT GATGGAGAT TGAATGAT 1560  
GCGTTAGGA GAGTGAAGC ACCCTTAG AGAGAGAT CATCTCTGC CATTCATCT 1620  
GCGTATGCT AACAGAGAG AGCTGATTA GGGAGTGTCC CAGCTAGCT GATCAAGGG 1680  
AATTTCAAG AGCGCTGGG CAGCGCTG NCCCGATG CAGCTGAG AGTGAAGAG 1680

5 CATTGGGAAA GTTCCACACC ACTGTGTAGA CCACTAGGTT CTCTGTTCCTT 1740  
CCTTTTCAAA TCCCNCAAGT TCTGTGTGG GAGAGCTGT AATTAGCTTA GTCCAGGTAC 1800  
CAGATCCGAG CTAGGGGGC AGCTGCTTGT GATTAATCCA AGAANAACCTG GGCACCACTA 1860  
TTTTTCCAAAT TATTAAGGACT GTGGCATAAA TTTTAAATAG AGTTATATGT AAACCAAGAT 1920  
TCTCCAGCTG CCAGGGGAAG AAGGTAGGGC TGGACTCCCT GCTGTGGGCC AGCCCTTGTT 1980  
AGGGGTGGT CTCTCACTGC AGCCAGACAG GATGATCTGT GGTTCGTGGG AGGGTAAGCT 2040  
GCCCTTGGC GAGTCTCTGA CCGAATTAAG AGTCCAAACC GCTCTCTTCC GTGTCTGTAG 2100  
AGATGGGTAA ATGGGTGATG GATGAGCAG ACTGAGAGCA CAGCAGATGA CTCATGTGTC 2160  
GAAGAGGGG GGAGATGCT GGGGTGGCTA GCTAATGTTC CCGCTTTTCA CGGATTTACA 2220  
GGAAATGGAG CCGAGCTTGG TCAATAGTT GTTTTCTTC CACTGTGCGA TGCATCTCTC 2280  
AGAAATTTTG AAGTCAAGCT GCACTTCTC GAAGACTTTC TTCTTGGGCT TBACTCTCTC 2340  
ATCTGTGGG CCGTTTTCAT AGCCCTTCAC AAACACTGTC TCACCGAGAG CAGAGCTCTC 2400  
CGAGGGTTC AGAGGTTTCA CTGGCGGTTT ATCCCTTTCA TAGAAGCAACA CAGAGCATG 2460  
CCTTGGGACT GCACTCTCTT CANTCTCTG GGTTCAGGT TGCACAGAC CACTACAGC 2520  
CTCTCTGCA GTTCTCTCTT GGGCAGCAG TGTACGAGC GGTCTACAC AGTCTGTGT 2580  
TCAGTCTCC CAGATCTAAT CTCTCTTACA TACAGGCTGT CTGATCTG GTCCTTCTCTC 2640  
ACAGTATGA TTTTCCCAAC ACGGATATCC AGCGGGGATG GATGACCTC CTCTGTCTCT 2700  
GATTTCTTGG CAGGCTTTG GCAATGGCT TCTGTCTTCA GGGATCTGG TAGGCGGCC 2760  
TGGCAGTTT TTTACGGCA GGGGTATTA ACTTTTCCCG GATTCGATCC AGCACTTGT 2820  
TCAGTGGAC TTCAACAGAA TTCTTCAGGT CTCACAGATG TACAACCTCA GCAAGAAAT 2880  
CCTTTTCCAG GTTCAAGTAA GCTGTGTAGG TTTTGTTCCT ACCCATTTTC TCATCTCTTA 2940  
GGATCAAAA CTGGACTTAA AGGGAAAA GGACATGCTT GATGAGAGAC AGAACCTCAT 3000  
TGTCTCTCAC ATTCTCTGGC TCACAGAGG CTTCTCTTCA TTTTCTTTC ACATCTCTCT 3060  
TCCATCAAG GAGATCAATC TTGAGCTCTT CTCTGAGCA GTCATTTTG CTCCTGTGTA 3120  
ATCTGTGAC CATGAGTTTC ATCAGATGA CCGGTTTTC ATAGCCAGAT GAGGAGGT 3180  
ACTTCTCTGC AAGGTGAAA ATCTTCTCTT GATCATGACC TCCAAATTTG GCACTTACTT 3240  
TTAAATATC TTCAATCAAA GCTCTCAGTC CCGGTTATTA GAGGCCACTC AGCAAAAGGT 3300  
GCTCCAGCTG CTTTACAGC TCAGGTGAG CTTTCTTGA ATCTGTGTGT GTGACCAAG 3360  
AGGAGGTCT GTACATCT ATGTGTACTT CTTTCTGAG CTGTTATCA GTCTTTTGA 3420  
TGAAGTTGAG CTCTCTCAG GGCACAAA TGTCTCTCAG CATTTCTTG CATCATCTCT 3480

5 CATGTAACCT GACTCGAGT TCTAGAGATT CCAATGGGGC TTTCAATGTTA TCCAGGTATG 3540  
CGTGGAGGTC GGCMAACAGA ATTGTACTT CACACCTGTC CTTTAAGAG TCTGCCAATCT 3600  
TTGACATGGG CACAAAGTAA GGCACATGTG GTTTGCCCCG GTTTCCTGTT CCCCAGTAAA 3660  
TTTTAAATTC CCGTCTCTTC AGTATCTCTT TCACTCTTTC TTCCCCCAAG ACCCTCTGCA 3720  
GTTTCCGGGT GATTAAGTGC AGTTTCTCTT CAGGGCTGGG AGCTTCCGCC ATGCTTCCGT 3780  
ACCCCTGCTT CCCCCGCTCA GCGCGCACCC AGAGCCCCCT CCGGGTCAAC CGTGCCTGCC 3840  
GCTTCCCGGG AACTGTCAAG CGAGT 3865

(2) INFORMATION FOR SEQ ID NO: 290:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1910 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

50 AGGAGAGCA GGAAGAGGGG TCTGCGCGCG GCGCTTACCC AGAAGCAGC GGAAGCAGC 60  
AGGAGTGGG CTGTCCCGCA GCGCAGCCCC GAGGAGAGCC CCCCCCGCC CCGGAGAGAC 120  
GGCTTTCCA GCGAGCCCCA CTCTTAGGAG GAGGGGAGGC GGGAAAGCAG CTCAGGCTTC 180  
ACCCAGCCGC CTGCCCCCAG CCGCGCACT CCCAGGCTCC TCGGACTCG CGGGTCTCTC 240  
CTGGAGTCT CCGAGGGGAC CGACTGTGCA GAGCGCATG AGTTGGTGT GGTCTTCTCTC 300  
TCAAGCTGC TGGCCCCCAT GGTCTTGGCC AGTGCAGCTG AAGAGGAGA GGAATATGAC 360  
CCTTTTCATT ATGATTACCA GACCTTAGG ATTTGGGGAC TGTGTGTGC TGTGTCTCTC 420  
TTCTGGTTG GGATCTCTCT TATCTTAAGT CCGAGGTGCA AGTCAAGTTT CAATCAGAG 480  
CCCCGGGCC CAGAGATGA GGAAGCCCCG GTGGAGAAC TCATCAGCC CAATGCAACA 540  
GAGCCCCAAG AAGCAGAGAA CTCAGTGTCA CCAATCAGGT GGAAGCTCTT GGAAGCTGAG 600  
CGCGCTCTT GAACCTTTGG ATCCAAATGT CGATGCTTAA GAAAACGGC CACTTCAGCA 660  
ACAGCCCTTT CCGCAGGAGA AGCAAGAAC TTGTGTGTCC CCAACCTTAT CCGCTCTTAC 720  
ACATTTCTC CAGCTGATGA TGCATCTTAC ACTTGCCTCC CCACTGAGC CTGCGGTCTT 780  
GCCACCTCC CGTGAATGT GTGTGTGTGT GTGTGTGTGT GACTGTGTGT GTTTCCTTAC 840  
TGTGTCTTCT GTGCTACTT GTTTGTGAT GGTATTTGTGT TTGTTAGTGA ACTGTGAGCT 900  
CGCTTTCCA GGAAGGGGCT GAGCCAGATG GGCATCTGCT CCTCTCTGCC CCGGTGGCCC 960



1020	TCGATCAAGCT TCTGCTCTTA GAGAGCTGCT TGTTCGCCGA GACCAAGCCC CTCCCTGAT
1080	TTAGAGATGC GTAGAGTAAAG AGCAAGGACA GTGTGTTCA GTGTCTTGG GACTGCGAA
1140	GGTTTGACAG AACTTGTGAT CATCTGAT GAGATCTTT GACTCTTTA ACAAATACCT
1200	TGCTTCTTAA TCCCACTGTA TCCCACTGT AGAGTCTTT AACACTGTA GATCAAAAC
1260	AAGAGCTGCG TGAAGCCAG GTTGACGTCA GCGAGCTAT GCGTTCCGT GATTATATTC
1320	TTCCCAAGGG CTTCACAGAG GATGCCCAAT CTGGCCCCC CATTACAGA GCGCCCCGGG
1380	ATTCCAGGCC CAGGGCTTCT ACTTGACCC TGGAGATGT GTCCCTCTCA TATCTGTCA
1440	GCATTAATCT CATTGAGCTCT GGAAGCTTAC CCGTTCCAC CTTCCTGCT TCTGAACCT
1500	CATCTACAG CCGAGCTGAT CCAAGTCAG ACTACATTC CTGCATTGG GTCTGTGGA
1560	GCATATGATT GAAAGACTGC TGTTCGTTG GAGCAGACAC ACCGAGATGG ATTGAGGAG
1620	ACAGAGGCC TTTCCTGTC TCCCTAACGC CCGTTAATG GCGAGCAGAG GCATCTCCG
1680	CATCTTTCG TCTGCTGTCT GATGTGAGA GCGGTAGAG AGTGTGTTG GAGATCTAC
1740	AGGCTTCGCG CAGCCTTGG GAAAGTGAAG AGGTGAAAG TCAATACAG ACTGAGACT
1800	CAGCCAGAT CCGGCCCTCT CTGTCTCTG TGTTCGCCG GAAACACAC AACCTTCCG
1860	CTGTACCCA TGTCTGTTCT CTGTATCTG ATCTATCTTC AACACAGCA GAAATTAAGA
1910	ATTAATATAT CTTGTCTTCA TAAAAAATA AAAAAAATA AGGGGAGGGG
35	(2) INFORMATION FOR SEQ ID NO: 291:
40	(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 3276 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
45	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 291:
60	GCACAGCTCG TTGAGATGCT GGTGAGGCT GCGGAGTACA CTGCACTGCG CACTTCGCG
120	ATCCAGAGCC AGGCTTGTTC GCGCCAGGCG TCCCTGCGCG TGGAGAGAG CGAGAGGAA
180	GCGCTTCCG GATTGTGAC GCTTCTGCG CCAAGCCTTG GAAAGAGCCA GCGCCGCGCC
240	AGTTCGCTCG TTTCACAGCG TGTAGGCTGT TAAAGCGGCG CCGCCAGAG CCGCGGCGCG
300	GAGAGGCGCG GCGCATGCT TCTGAGAGCG ATTCAAAAG TGAATGACTA TCACAGCTA
360	TTCTCAACA GAGAGAGCT CCGAATGCT TAAATTGTA TGAAGCATC AATGAGACA
420	AGAGTGTGCT GTCTTGTCT CAGCCAGAGA TGAATGATT GCAATGTTC CAGAGTGACA
480	CAGTGTCTT GAAAGGAGAG AAGAGAGAG AAGCTTTTTC CATCTCTCTT TCTGATGATA

540	CTTGTCTTGA TGAAGATATT CCAATGAAAT GAGTGTTCG GAAATACCT CTGTACGCC
600	TAGAGATGCT CATTAGCATC CAGCAGTCC CTGATGTGAA GTAGCGAAA CATTACATG
660	TGCTGCCAT TGAATGACA GTGAAAGCA TTACTGTGTA TCTGTCCAG GTATACCTTA
720	AGCGTACTT CTTGAGAGC TATTCAGCA TCCGAAAGG AATCATTTT CTGTGCGCTG
780	GTGGAGATGG TCGTGTGAG TTCAAAATG TGAAGAGCA TCTTACGCTT TATTCATATG
840	TTCTCTGAGA CAGATGATC CACTGAGAG GAGAGCTTAT CAAAGCAGAG GATGAGAGAG
900	AGCTCTTGA TGAATGAGG TATGATGACA TTGTGTGCTC CAGAGACAG CTACGCTGGA
960	TAAAGAGAT GTTGAATCT GCGGTAGAC ATCTGAGCT CTTAAAGCA ATTGTGTGA
1020	AGCCTTCAG AAGATTCCTG CTTACAGAC CTCTGTAAC AAGAAAGACC CTGATGTCTC
1080	GAGCTTACG AATAGAGACT GAGCCTTCT TCTTCTGAT CATTGCTCT GAGATCATGA
1140	CGAATGTGC TGTGAGATCT GAGAGCAC TTCTGTAAAC CTTTGAGAG GCTGAGAGGA
1200	ATCTCTTCC CATCATCTTC ATTGATGAG TGAATGCTAT CCGTCCCAA AAGAGAAAA
1260	CTCATGAGCA GATGAGAGCG CCGATTGAT CAGATTTTGT GAGCTGTGAG GATGCGCTTA
1320	AGCAAGGCC AATATGATT GTTATCCAG CAAACAGAG ACCCAAGAC ATTGAGCCAG
1380	CTTACAGCG ATTGTGTGCG TTTCACAGCG AGGTAAATAT TGAATTTCT GATCTACAG
1440	GAGCTTGAAG GATCTTTCAG ATCCATACCA AAGAACATGA GCTGTGAGAT CATGTGAGAC
1500	TGCACATGA GCAATGAGA CTACAGGCGA TGTGATGCT GACTTACAG CCGTGTGCTC
1560	AGAGCTGCT CTGCAGCCA TCCCAAGAA GATGATCTC ATTGACTTGG AAGATGAGAC
1620	CATTATGCC GAGGTACAGA ACTCTTACG AGTTTCTAT GATGACTTCC GGTGCGCTTT
1680	GAGCAGATAT AAGCATTCAG CACTGCGGGA AAGCTGTGTA GAGCTGCAC AGGTAACTCG
1740	GGAAGCATC GCGGCGCTTGG AAGATGTCA ACTGTAGCTA CAGAGCTGCG TCCAGTATCC
1800	TGTGAGACG CCAAGCAAT TCTGAAATTT TGCATGACA CTTTCAGAG GATGTCTGTT
1860	CTATGAGCT CCTGCGCTG GAAATACCT GTTGGCGAAA GCGATTCATA ATCAATGCA
1920	GCGCACTTCT ATCTGCATCA AAGTTCCTGA GCTGTGACC ATGTGTTTT GAGAGTCTGA
1980	GCGCAATTC AAGAAATTT TTGACAGGC CCGCAAGCT GCGCCCTGCG TGCATTTCT
2040	TGATGAGCT GATGTGATTT CGAAGCTCG TGAAGTAACT ATTGAGATG GTGTGCGGCG
2100	TCTGACAGCA GTATCATCAC AAGTCTGAG AAGAAATGAT GCGATGTCCA GAAATTAATA
2160	TGTGTGATC ATTGAGCTTA CCAAGCGCC TGAATCATTT GATCTGACA TCTCTACAGC
2220	TGCGCTGTT GATTCAGCTA TCTACATCCC AACTTGTGAT GAGAGTCCC GTGTGCGAT
2280	CTTCAGAGCT AAGCTGAGCA AGTCCCGAT TGCAGAGAT GTGAGCTGCG AGTTCCTGCG

521

522

TAATAATCACT AATGGCTTCT CTGAGCTGA CCTGACAGAG ATTTCGCAAC GTGCTTCGAA 2340  
GCTGGCCATC CCGTAATCCA TCGAGAGTGA GATTAGGCGA GACAGAGAGA GCGAGACAAA 2400  
CCCATCAGCC ATCGAGGTAG AAGAGGATGA TCCAGTGCCT GAGATCCGTC GAGATCAGCTT 2460  
TCGAGAGGCC ATGCGCTTTC GCGCGCTTTC TGTCACTGAC AATGACATTC GGAAGTATCA 2520  
GATGTTTGGC CAGACCTTTC AGCAGAGTGC GCGCTTTGGC AGCTTCAGAT TCCCTTCAGG 2580  
GACCAAGGTT GAGCTGGCC CAGTCAAGG CAGTGGAGGC GGCACAGGTG GCGATGTATA 2640  
CACAGAGAC AATGATGATG ACTGTATGCG CTAACTGTGT GTGGCCAGCG TCGAGTGAGC 2700  
TGGCTTCCTT GAGCTTGTAT CCTTGGGGT GCGGGGGCTT GCGCAGGAGA GGCACAGCG 2760  
GTGGGCCAC AGCTGCTCC ATTCCTCAGT CTGACAGTT CAGCTACAGT CTGACTCTGG 2820  
ACAGGGGGTT TCTGTTGCAA AATACAAA CAAGAGCAT AATTAAGG CAGTTTTCAT 2880  
TTGGTAGCGG GAGAGTGAT TACCAACAGG GAATGGGCC TTGGGCTATG CCATTTCTGT 2940  
TGTAGTTTGG GCGAGTCCAG GCGACTGTG TGGGTTGTGA ACCAAGGCAC TACTGCCACC 3000  
TCCACAGTA AAGCATCTGC ACTGACTCA ATGCTCCCG AGCCCTGCTT TCCCCATC 3060  
CAACTGGGTT AGTGGGTAG GCGCCACAGT TCGTGCATGT TTATATAGAG AGTAGGTGA 3120  
TTTATTTTAC ATGCTTTTGA GTTAATGTTG GAAACTAAT CACAACAGT TTCTAAGCA 3180  
AATAATGACA TGTGTTAATA GGCATATAA GTTTGGGTCT AATGGGHA AAAAAAAAA 3240  
AATAAGGG GCGCTTCTTA AAGNCCANN CTTCGT 3276

(2) INFORMATION FOR SEQ ID NO: 292:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1695 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

TTGCATGCT TGAATTCGCC TCTCAGCC AGCTAGAG AGAGGTTCG TAGTCCAGA 60  
GCTGAGGCAG GAGCGGCGAG TTTCTGGCGG GTGAGGGCGG AGCTGAAGTG ACAGCGGAGG 120  
CGGAAGCAC GTTGTGGTGG CCGAGAGAG GCGCTGGCCC CAGGAGGAGG AGGAACCTT 180  
TCCGAGAAA CAGCAGCAG CTGAGCTCT GTGACAGAGG GGAACAGAT GCGCGGCGCG 240  
AAGGAGCCT CTGGGTGAGG ACCCAACTGG GCGTCCGCGC GCTGCTGCTG CTGACCATGG 300  
CCTTGGCGCG AGGTTGCGGG ACCGCTTCGG CTGAAGCATT TGAATCGTTC TTGGGTGATA 360

GCGCTCTTTC CGACCGGGCC TGTGAGTTGA CCTACCCCTT GCACAGCTAC CTTAGAGAGG 420  
AGAGTTGTA CCGATGTCAG AGAGTTGCA GCGCTTTTTC AATTTGTGAG TTTGTGAGTG 480  
ATGGAATGTA CTTAAATGCA ACTAAATTCG AATGTGAATC TGCATGTACA GAAGCATATT 540  
CCCAATGTA TGAACATAT GCTTGCATC TTGTTGGCA GAATCAGCTG CCAATCGCTG 600  
ACTGAGCA AGACAGACTT ATGTCCCTGA TCCCAAAAT GCACCTACTC TTTCTCTTAA 660  
CTCTGTGAG GTCAATCTGG AGTCAATGA TGGACTCCGC ACAGAGCTTC ATAACCTCTT 720  
CATGCACTT TTAATCTGAA CCGGATGAG GAAATATAG TATATTCAG TCTAAGCCAG 780  
AATCCAGTA CCGACAGAT TTGAGCAGG AGCTACAAA TTTCAGAGAA TCACTCTTAA 840  
GCAAAATGTC CTATCTGCAA ATGAGAAATT CACAAGCCA CAGGAATTTT CTTGAAGATG 900  
GAGAAATGTA TGGCTTTTTA AGATGCCCTT CTCTTAATC TGGGTGANT TTAACCTACA 960  
CTCTGTGCT CTGGTGATG GTATGCTTTT GGAATTTGTT TGCACCTGTT GCTACAGCTG 1020  
TGGCAGTA TGTTCCTCT GAGAGCTGA GTATCTATGG TGACTTGGAG TTTATGAATG 1080  
ACAAAAGCT AAACAGATAT CAGCTTCTTT CTCTTGTGT TGTAGATCT AAAACTGAG 1140  
ATCATGAGA AGCAGGGCTT CTACTACAA AAGTGATCT TCTCATCTCT GAATTTTAG 1200  
CATTTTCTT TTAAGAGCA AGTGTATAG ACATCTAAA TTCCACTCTC CATAGAGCTT 1260  
TTAAATGCT TCAATGAT ATAGGCTTAA AGAATCACT ATAAATGCA AATAAGTTA 1320  
CTCAATCTG TGAAGACTGT ATTGCTATA ACTTTATGCG TAATGTTTTT GTAGTAATTT 1380  
AAGAGTGA TGTTTGGAT TGTATATTA TTTTACTAAT ATCTGTAGT ATTTGTTTTT 1440  
TTGCTTGTGT TAATGTTTTT TTCCCTTTC TTAGCTATGA GCTGATCAIT GCTCCTTCTC 1500  
ACCTCTGCC ATCATACTGT CAGTTACCTT AGTTAACAG CTGAATATTT AGTAAATG 1560  
ATGCTTCTGC TCAGGATGG CCGCAGATC TGTATTTTGA AATTAGCAG GAATGACCT 1620  
TTAATGACAC TACATTTTCA GGAAGTAAA TCAATTAAAT TTTATTTGAA TAATTAAAA 1680  
AAAAAAAAA AATCT 1695

(2) INFORMATION FOR SEQ ID NO: 293:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1501 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

CACCTTCAGC AGTCTTTTGC TGTCTTCTT TCTACCTCAA ATAGCCCGAG GAGTGGGCTT 60

	ТАСГТССАА ТАРСААГАТ СГСААСТГТ ТСТСВВВВВВ ТВВВВВВВВВ ВАРВВВВВВ	120
	АТСГТТТТТ ВАРТСТСВВВВВВ ТТАТТТТВВВ ТВВВВВВВВВ АВАААААААА ВВВВВВВВВВ	180
5	ТСТТТАТТТ СГВАВВВВВВВВ ВТАААААААА СГВТАТТААА АСГТТТТТТ ССГТААСТТ	240
	ТСТТТТТТТ ВВВВВВВВВВ АТТТТАААААА СГТТТТТТТТ ТТАААААААА	300
10	ВВВВВВВВВВ СГТТАВВВВВВ ВВВВВВВВВВ ВВВВВВВВВВ ВВВВВВВВВВ	360
	ТВВВВВВВВВ СГТТТТТТТТ АВАААААААА ВВВВВВВВВВ АВАААААААА	420
	ТСТТТТТТТТ АТАААААААА ВВВВВВВВВВ ССГТААСТТТ ВВВВВВВВВВ	480
15	ТТАТТААААА АТАААААААА ТТАТТААААА ТСТТТТТТТТ АТТТТААААА	540
	ТААТАААААА ТТАААААААА АВАААААААА АТАААААААА ТТАААААААА	600
20	ТТАТТТТТТТ СГТТААААААА ТВВВВВВВВВ ВВВВВВВВВВ ВВВВВВВВВВ	660
	ТТАААААААА АСГТТААААА СГТТТТТТТТ ТТАААААААА ТВВВВВВВВВВ	720
25	АСТТТТТТТ ТТТТТТТТТ СГТТАТТАТТТ ТТТТТТТТТ СГТТТТТТТТ АТАААААААА	780
	ТААААААААА АВАААААААА АВАААААААА СГТТААААААА АТАААААААА	840
30	СГАТТААААА ТВВВВВВВВВ ТВВВВВВВВВ ТСТТТТТТТ ТТТАААААААА	900
	ТААТТААААА АААААААААА АВАААААААА АТАААААААА ТВАААААААА	960
	ТААТТТТТТТ ТААТТТТТТТ АТАААААААА ТВВВВВВВВВ ТТАААААААА	1020
35	ТТАААААААА АТАААААААА АТАААААААА АВАААААААА ТТТТТТТТТТ	1080
	АТАААААААА АСГТТААААА АВАААААААА ВВВВВВВВВВ СГТТААААААА	1140
	АВАААААААА АВАААААААА СГТТТТТТТТ ТСТТТТТТТ АТАААААААА	1200
40	ТААААААААА АВАААААААА СГТТААААААА ТТТТТТТТТ АТАААААААА	1260
	СГТТААААААА СГТТААААААА ТТАААААААА ТТАААААААА СГТТААААААА	1320
45	ВВВВВВВВВВ СГТТТТТТТТТ ТВАААААААА СГТТТТТТТТТ ТТАААААААА	1380
	ТААТТТТТТТ ТТАААААААА ТТТТТТТТТТ СГТТТТТТТТТ АВАААААААА	1440
50	АТАААААААА АТАААААААА СГТТААААААА АВАААААААА СГТТААААААА	1500
	А	1501

(2) INFORMATION FOR SEQ ID NO: 294:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2683 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: Linear

(E1) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

5	ТААТТТТТТ ССВВВВВВВВ ВВВВВВВВВВ ВВВВВВВВВВ ВВВВВВВВВВ	60
	АТАААААААА ССВВВВВВВВ ССВВВВВВВВ ССВВВВВВВВ ССВВВВВВВВ	120
10	СГТТТТТТТТ ВВВВВВВВВВ ВВВВВВВВВВ ВВВВВВВВВВ ВВВВВВВВВВ	180
	ТТТТТТТТТТ ТААТТТТТТТ АСГТТААААА АТАААААААА СГТТААААААА	240
15	СГТТААААААААА АТАААААААА АВАААААААА АВАААААААА	300
	СГТТААААААААА АТАААААААА АВАААААААА АВАААААААА	360
20	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	420
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	480
25	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	540
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	600
30	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	660
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	720
35	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	780
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	840
40	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	900
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	960
45	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1020
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1080
50	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1140
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1200
55	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1260
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1320
60	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1380
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1440
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1500
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1560
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1620
	АВАААААААА АТАААААААА АВАААААААА АВАААААААА	1680

5 GGAGCTTGCA TTTCAGTTTG TTGTGATGC CCGCGAAMAA GTGGGGSCA AACAGCTGA 1740  
AGATGCTGCC AAAAAAGTAA CCATGGGCA GTGATGATGG AGCTAGGCC TCCAGCTCCA 1800  
CAGCTCAAGA CGAAGATGAC GTTCTCATAG TTGATTTGGA TGAAGAGAT TCTTCAATA 1860  
ATGCGAGCT CATGAGGAG AGAGAGCCG CAGAGGAAA TTAGATGAGA AAGAGATCT 1920  
CACTGCAAG AGCTCAGTA TAGACAGAA CGAGAGCTT GATGATGCA TAGCATAGA 1980  
TTGACAGAA ATGCTCTTAA ACAGAGCCT CTTACTATT AGTTTATCTG GGCAGACCA 2040  
GATTGTATG TCTTCTGTT CAAAGGAAA AAATGACAG CAGTACATG AAATGATTC 2100  
TCTTCCCTTT GAAGCATTG ATTTTGTAG AACTGTAGA CAGTTGCG TATGCTGTAT 2160  
TGAAGTAGG AATATAGTTT TAAAAACCT TTGAACAAAG TGTGTGCATA ACCAGTCATG 2220  
AGATAAACA ACAAAATGCA TTTTGCCTT TTATGTAAA TACCTTAGG TATCATTAAT 2280  
AGTTTCAAAA TATTTGTGTT TATTAAGTT GATACCTGCT TATTAATATT ATGCTTTAT 2340  
TTTGTGCTAG AAGAGAAATT ATTTTATGCC TAGATCTAAC CATTTTCATA CTCTTAAGT 2400  
ATTGAACAG ATTCAAGAA GTATCGAGTG CTATGCAATG AAATCTTATT TTAAATGTTA 2460  
GATGGACTA TGTATATTA TGTAAACAA TGTAAATTA CTCAGTTTT CAGTTTGTAC 2520  
CGCTGCTAT GTCTGTGTA GAAGCAATT TTTGTGTATT GTTACAGTTT CAGTTTATT 2580  
ATATTCGATG TTTTGTAAA CTCAAATAC GACTATACTT ATGACCAAA TAAATGGCA 2640  
TGCATTCGK TTAAMAAAN NACAGAAA AAAAAACA AGA 2683

35

## (2) INFORMATION FOR SEQ ID NO: 295:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1454 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

45

50 GGACTCGGGG TGGCTCTTAG GGGCAGGAT AGGGCTGGG AGCGCGGGC TGTGGCCGTG 60  
ACCAGCCCT TCTGTGCGAG GTTCCACCC GATCGAGTG GTCAATGCT TGAAGGGA 120  
CAGTACCTG AGCCACTGCT TCTCCAGCA CTCATGTC GTCTGTCT CTCTGAAGG 180  
CAGCCCTTG CCGAGCCTG TTGACAGGA CTTCCTACT CAGTTTGGG ACAAGACAC 240  
AGGGAGATG GAGAACTAG ACCTGATCCA CTCTAGTGG GTCAAGTTTA CTTACCCGAG 300  
TGAGGAGAG ATTGGGACC TGAAGTTTAC TGTGGCCAA AAGATGGCT AGCCAGAGA 360  
GGCCCCAGC CTCAGCATTC TCTGTACT GTAGGCTTC CAGTGGGCA TGGCAGCCCC 420

60

5 TGGTCTGC AGGGGCCCC TCGCCCCAA GACACTCTG CTCACAGCT CGAGATCTT 480  
CTCTCTGAT GAGGACTGTG TCCACTACCC ACTGCGCGAG TTGCGCAAG AGCGCGGCA 540  
GAGAGACAGG TACCGGCTGG AGATGGCGG CCGCGTCCG GACCTGACC GATGCTCAT 600  
GGCTACCGAG ACCTACCGCG AGCCCTACCC CTCTGTTTCG ATGAGTGCA AGTCAATAC 660  
CTATGAGGCA GTGTCACTT GAGCAGCTT GGGAGATGC CAGGTGGCC GCTTAGGCC 720  
AGCAGGGCC GTGAGTTCCA GTGGAGGTG TTGTGCCCA GTCTGAGAG CAGAGAGAAG 780  
CTCATCTGC TGTGGCTCG CCAGTGGAG GCCCTGTGTG GCTGAGCTG CTTGTGAGC 840  
TCACCGGCTA GCGCAGGCA CAGCAGCCT GTCTGTGCA GCTGAGGCC TACTGGGCA 900  
GGCAGCAGG CTTTGTGTT CTCTAATAAT GTTTATCTT CCGTTTGTG CTTTAAATTG 960  
ACTGTCTCG CAGAAATGTG AACATGTGTG TGTGTGTGT TAAATCTTC TCATGTGG 1020  
AGTGAGAAAT CCGGGCCCC CAGGGCTGTT CCGTGTCTG TCAGCTGCC ACAGGTGTA 1080  
CAGCGCTGCA CACCAATGTC GTGTCTGCTG TTGTGGACC GTTGTAAAC GGTGACACTG 1140  
TGGTCTGAC TTTTCTCTT ACAGTCTCTT TCTGAAAGTG TCGATGCCG TCGTTGTG 1200  
CTGTCTGCT TCGTGTGCT GTTCTGTG GCACTTCTT GCTATCTGT AGGCTGTAG 1260  
CAGATGAC ATTCGAGCT CCGACCCAT ATTGTCTTC AAGTGGAGG TCTCCCTGA 1320  
TCCAGACAG TGGAGAGCC CCGGGGCA GGGAGCCTG AGCTGCCAG ACCAAGCTG 1380  
ATTCTGCTG CCGTATCTT CTATTCGAT AAGCAGAT TTGACACCG AAAAAAAAAA 1440  
AAAAAAAAA AACN 1454

35

## (2) INFORMATION FOR SEQ ID NO: 296:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 828 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

45

50 ACCCTGGAT GCGCCGAAA CAGATACCA GCGAGCTTAC ACAGGATTA ACTCTCTCA 60  
ATGAGGAGA ATCATTTACA ACTGAGCAG ACATCTATAT GATCATTTAA GGAAGTGT 120  
CCCTTATGT TTAGCAAGTA TAATGGCTA ACTCTAATAT CCGAATGAAT AGTCTAGGC 180  
TGGACGAA TGGGCTCAA TTAGCAGAT AAGGATCA GTCCAGTAA ATGATTCAT 240  
AGACTCATCT AGCAGCACT ACCATAGCA CTATGTTAG AGCTGAGG CCGCAAGTA 300

60

360 GAAAGATGTC ATTAATGTCG CTCCTGTGTA GGTCCAGAGA GAATTCGAC ACAGACATA  
420 CAGTTAAC TCAGTCGAC CTGCAAGTAA TACGACAA ACCTGAAATA ATACCTATG  
5 480 AAGGAGGAGG GCTGAGGTC GGCCTGAG GATGATGAA ATTGATGAG AGATGAGAA  
540 AGACAGAGG ATCCAGAGTC AGAGAGCAT GAAAAATGAG CAGGGGCTG GATCGATGAG  
600 GTGTATCAG AGCAGCTTTC CAGATCGAC ATGATGCTC AGATGCTTT GCTATGCTGT  
660 GCGAGATGT CCGAGCGAGA TGTGTGCCCC GAGCCCCAT CCAATTACAT GTCTGTCAAT  
700 GCGGACCTCA AAGGTACAT GTTGTGAAA GCTTTCCTTC GGTATCGAGA ATCGAAATTA  
15 780 ATCGAGGATC TTTCAGACAT GGTGTTTTTT CCTCTTGGGT CTTCTATGTA CTAAAGCTCA  
828 TCTCATGAG CTTTACAGCA TTAATAATTA TTGTTTTTTC TGCATACA

20 (2) INFORMATION FOR SEQ ID NO: 297:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2416 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: Linear

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

60 TCAATTCGA TTACATGAGA TCGACCATTC TCAATTCAGA TTGTTCAGGC TCTCAGCTTT  
120 AACAAACCT ACTATACAGA TGTATCTTCA ACAGCGAG TCATGAATTA GCGACATTT  
35 180 TTTTTCCTTT GAGAAAGAG TCGACTGGGG CAGCACTTTT AGTCGAGGG GAGCTAGTGG  
240 AAATCTAGAC AATGAGAGTC ATCGATAGCA GCTTTCCTTC AAATGTGTGA CTCTCCAGGG  
40 300 GCTAACTGAC TCTTAGCTTA GAATTAATCT TTACATGAGA TCGACAGAT AAGTGGGTTA  
360 ATCGATAGCA TCTGTATACT AGTTATATAG CTTCAGACA TCGAGAGAC ATCGAAAGAG  
420 GATGAGACA ACCCGAAGGA TATCGAAGAA GCGCATATAG AACCCCTTC CTTCTGGGAG  
45 480 GAGAGAGAG CAGACAGAG GAGAGAGTC AAAGCATTTA GATGTTTAA GAGAGAGAG  
540 GCGAAGCTTT GACAGTCTT TGCCTTTTGC CAGTTTCAG CAGTTCCTGG CTCTTGCAA  
600 CCTTAGGTC AGTACATTT GCGAATTCG AGATTTGCC CAGATATGA GAGAGAGAG  
660 GACTTTGGGG GGTGGGAGAG GGTGTGCTGT GTTTTAAAG GATAGTTAG CTGTTCGAC  
700 TGTATTAGA TAGAGAAAA AATATGAGGG CAGGTGAC ACAGACATA AATTTGTG  
55 780 TTTTATTTT GTGATGCTGT TGAATATGTT TACACATTC TAGATATAC AGTGAACCT  
840 GATGTCTGT TCGATTAAC ACTATATTTT TTTCGAAATG TTACTGTCA AAAGCTGTT  
900 CCGTCCCTTT CTTTTCCTTA TGTACTCTCT TCAATCTGAC TTACTATGAC ACCGAGCAA

960 TACCCATCCA AAGACTGAGC GATGAAAGAG GGCCTTTTCC AAGTACGCTC TGGGTGCTCT  
1020 AAGCGAGCT GTGCTCTCTG GTTTATGAGG TGTATATAG TGCCTGGAC CTTTCTTTTC  
5 1080 AATGTAGCT AACAGAGAG ACTGAGAGCA GTTATCGAT TCTGTATCA GATCGACTAG  
1140 GATGAGAGG CCGAGGAGAT GAGGGGGTTC CTGAAAGTTT GCAATGGCT GTAGTACTG  
10 1200 AATCTTTTC GATGTAGAG AATCTGAC CAGTTAGAT TTACATGAGA AAAGGTAGA  
1260 TCAATTCAG CAGGTATGAG CTGTGTGCGA GGAATGTGTA AGCGAGAGAA CTGTCTATCA  
1320 GCGAAACCTC AAGCTTCCG TTTTAGAGAG CAGCAGAGAG AGGAGCTTGG TCTCTCTGTT  
15 1380 CTGTACTTTC TGTACATGCG GGTGACCTGA GAGCTGCACT CAGCTGGGG AGCGAAAAAG  
1440 GAGAGGTGAT TCTCTTTTCT CTCCCGACCC CTTGCCCTTT GTTACAGACA CAGATTTCGC  
1500 AAGGGGTACA TGAATTTCTG AATTTTAA AATGTTTTT GATTGTGTT TTCTGGGAGC  
20 1560 TGAATAGTGC TTTCAGAGAT GTTCATAGCC CTGCAAGACT CCGAGCTTAG TCAATTTCTT  
1620 GTAAATTTCT GTTCAGATA TTGTGTGTG TGCCTGTTTT GCGAGCTGAT TTTCGCTGTA  
25 1680 TTAATATGTC AGTATGATAT TGAATCTCTT TTTCCTTAA GGAATATGAA TTGTTTTTCT  
1740 TGTGTATAT TCTGCTTGTG AATACATGAG GCAAACTGAG GGTGTAGAG CTGATAGTAG  
1800 GGTGTGAAT CAGAGAGAGA GCGGTGAG TGTACTGTGTC CTTGACAGGC TGAATACCT  
30 1860 GATGTCTGA GCTTTCATCT CCAATCTTT GCAAGCTCA AATGTGACA GAACTTTCGC  
1920 TTTTCTGCC ACTGCCATG GAGAGAGAG GAGCATGCTT ACATCGAACA TGTCTGTGCT  
35 1980 CCGAGCCCTC CCAATGCAAT GCGAAAGAG GTTACTTTGG GGAATGGGG GATTAGATGAG  
2040 GATGCTGTG TATGAGACA CTAGACCTTC TCTCTGCCCT CCGTCTGCC CCGAGATCA  
40 2100 CTTCCTAGGA GAGCGAGCT GCTTGCAGAG GATGCTGTTT CCGTCTGAC TCGAGAGAG  
2160 GATCCAGAGG CTTTGTAGCA GTCTGAAAT CCGTTCGCTT CTGTATAGA ATCAATATGT  
2220 AATGACCTT TCAAGCATTA CAGATATCAT TGGGTATCT GTATATATAT CAGTACTG  
45 2280 GAATGTGAA GGTGAGAGTA GTTAATCTTT TACCTTTAT TGTATATTT TTGTATATAC  
2340 TGTCAATGAG TATTAATAT TGAATGATTT GATGCTTTT AATAACTGA AGCGAGAGAG  
2400 GAGCACTGCG TCGAGCTTTC CAGTGTGTA TGAAGCACTG TTTTTAAATA AAGAGAGAA  
2416 KCAAGAAAA AAAAAA

55 (2) INFORMATION FOR SEQ ID NO: 298:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 545 base pairs

60

529

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

5 GAATTCGGCA CGAGCCATCC YTGCCCTCTC CTTCATCTTT ACAGTCACTT TGTTCCTGT 60  
10 TTCTGACTCA GCACTACTCT GATTTGTGGC CAAGGAGTGA CTTATTTCTT CTCAGGAGGG 120  
CAAAAATGTG GAATAGTGTG TGTCCATGCC TCTCTCTCATG GCTTACCACC TCTGCCACCG 180  
TGTATATCA GTAAACACCA GGAGAGAGCC TGCTGGAACT GACTCTCTGG AACTCCCTGG 240  
15 ATGTTTGTGT GCAGGAATGT AGTAGGCAAT CACGTGTGTG GTTGATCTG GGGCTTCTGT 300  
ATGTGAGTAG AGAGTAAAA GGSACCACAT TCTTTGACCT YTGCGGAAT CATCTCAAAA 360  
GAAATGTTT CCAAGATGCT TCTGAAGATT GCTTAAATAT AGCCGGTTTC CACCCCTGTG 420  
20 AATGCATCCA TTCTAGATAG CTCCTTCACC AGGACACAGG AACTGATTTA CAGAGTGAC 480  
ATGAATACAT TCCATCCAG AATTTCGANT ACTTCAATTT NAAATTTCTAC CTATTAAAA 540  
25 NAAAA 545

(2) INFORMATION FOR SEQ ID NO: 299:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1530 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

40 GGCCTCTGCTG GGCATCATAC TTGTCACTCG GTAAACAGTT TGGCCACTTA CCGCAGATGA 60  
AGCTGCTTGC CAGGCTCTC CCGCTCTGTG AGTTTGGGAG CCAGGCATCT TCACAGAGGC 120  
45 TGTGTGCTGG CCAGGAGATG GTGGGGCCCC GCGAGGGGTG CTCGGCTCCC GTTCAGGTGG 180  
TTGGGCCCCAG GCGTGATCTC CCACCTGTG GAGCTTCAT TACTGGAGG ATCATCGGGC 240  
CAGATGATGC CAAGTGCGCC GGCATGTGTC ACGGGGGGAC CATCTCTGAG ATGATCGAGG 300  
50 AGGCGAGGGC CATCATGAGC ACCCGGCATT GCACAGCCA GAAGGGGGAG GCTGTGTGG 360  
CGGCTCTGGC TGTGTGGAG GCGACCGACT TCTGTCTTCC CATGTGCATC GTTGAGGTGG 420  
GCGATGTGAG CCGCGAGATC ACTTACACTC CTAAGCATC TGTGGAGGTG CAGGTCAAGG 480  
55 TGAATTCGCA AATACATCTC AAGGTGCGCA AATAGCTGAC CAATAGAGGC ACGCTGTGTT 540  
ATGTGCCCCCT GTGCTCGAAG ATGTGAGCA AGTCTCTGCA GGTGGCTCTT GTTGTGTTT 600  
60 CCGCGCANGA CGAGGAGGAG GAGGGCCGGA AGCGGTATGA AGCCGAGAG CTCGAGGCCA 660

530

5 TGGAGACCAA GTGGAGGAC GGGGACATCG TCCAGCGAGT CCTCAACCCA GAGCCGAAACA 720  
CTGTCACTTA CAGCAGTCC AGCTTGATCC ACTTGTGGG GCTTTCAGAC TGCACCTGTC 780  
ACGGCTTGTG GCACGGAGGT GTACCATGGA AGCTCATGGA TGAAGTCCGC GGGATCTGTG 840  
CTGCACCCA CTCGAGAGCC AATATGTGCA CAGCTTCCGT GGAGCGCATT AATTTCATCG 900  
10 ACAGATCAG AATAGGCTGC GTCAATACCA TCTCGGAGCG CATGACCTTC ACGACCAATA 960  
AGTCCATGGA GATCGAGGTG TTGTGTGAGC CCGACCTGTT TGTGAGAGC TCTCAGAGCC 1020  
GCTACCCGGC CCGCAGTGGC TTCTTCACTT ACGTGTGGT GAGCCAGGAA GCGAGGTGTC 1080  
15 TGCCTGTGCC CAGCTGTGTG CCGGAGACCG AGGACGAGAA GAAGCGCTTT GAGGAGGCCA 1140  
AAGGCGGTTA CTTGCAGATG AAGCGAAGC GACACGGCCA CCGCGAGCCT CAGCCCTAGA 1200  
CTCCCTCTTC CTCGACTGCG TGCCTGAGT AGCATATGCA ACGGGGCCAG TGTCCAGTCA 1260  
CTTAGAGGTT CCCCCCTGG CCAAAAGCC AATTTCACAT GAGAGCTGTT GTTGTCTGAA 1320  
20 GTTTTCTGAT CACAGTGTTA ACTGTACTC TCTCTTCCA ACTTACACAC CAAGCTTTTA 1380  
TTTATATCAT TCAATATCA ATCTACACA GTGTGTGTC GAGGCGGGG AGGCTTGTGG 1440  
CAGAAACCTT CCGGATGCTT TCCGAGACG CTGTAGGTTA TGGGAGAAC CCGACGACAC 1500  
30 TATATAGCT GATGCTTGGC TGGGAGAA 1530

(2) INFORMATION FOR SEQ ID NO: 300:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 997 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

45 AGGTAGTGGAG AGACACATTA CACTTACCA ACAGAGAGAA GGAATCTCCC CTTTATATT 60  
TAACTATGTT TACAGGGAAT GCTTACATG TGGCTTCCCG AGNATTTGCT CCAACATGTT 120  
TTGAGAGACC CTAAATGCCA ACACTGATT GAATGGGTAA AAGACACTTA TAGCCGAGAT 180  
GAACACTCT GGGCACCCCT TCAGGTGCA CCGTGGATGC CTGGCTCTGT TCCCAACCAAC 240  
CCGAGTAGCG ACATCTTCAG ACATGACTTC TATTTGCCAG CTGTCTCAGT GGCAGGTTCA 300  
55 TGAAGGAGAC ATCGATAGG GTCTCTCTTA TGTCTCCCTGC TCTGGATTC ACCAGCGGGC 360  
TATCTCGTT TATGGGGCTG GGGACTTGAA TTGATGCTT CAATACCATC ACCTGTGTCG 420  
CAACAGTTT GAGCCGAGG TAGTGTATTA TGTCTCTCAG TCTTAGAAG AATACCTAGC 480

TTTAAAGCC ATCTATGGA CTGAACTTGG AAGACACTA TGAAGAGCTT GCTACCTTGG 540  
GGGAGAGAC ATGTGAAAC ATGCTGAGA CTTCCTGGA CAGTGTGGT GGGAGACAG 600  
GAGTTGGA TTCTGGGANT CCTTGAAGAT AAGAGAGCTG KATTTAAATT GTGGTAAGT 660  
AAGCTTTTG CCTTGGAAT TGTGAGCTGG GTGATCTGG CTGTGCTCTT GAGGCTTAC 720  
GCTTAAGTT CCTCACTTA CTTCCTCACT AAGTAGAAT GAGAACTCT GTGATGAGA 780  
GATGAGGA GGGATTTGTG GTAGAGACT TGAATTCACT TGAATGCTT CTGATAGCTT 840  
TTCAATCTG TGAAGCTGCC GTTCCTAAAT ATTCAGCTT TGTAAAGCTT GAGAGAACT 900  
TTGATGAAA GAGAACTTC CCTCTGTAC TGTAACTTA AAAATTAAT GCTCTGATT 960  
CAAAATGAG AAAAATAAA AAGAAAAA AACTGGA 997

## (2) INFORMATION FOR SEQ ID NO: 301:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2345 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (2) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

TTGAGCGGA CGCTGAGGC CGGAGAGAA ACTGAGAGC GAAAGTACC GGGAGCCAG 60  
CAATTCAAT CTGCTCGATA GACTGTGTC AACACACAA TGTGGCTGTC AAGGCTGTG 120  
TGTCTCGGA CACTACCTTC TAGGGTTTTC CAGCCAGCTT TCAACAGAGC CTCCCTCTT 180  
GTGAAGAAAT CGATGAGGA GAATCAATGG CTGTAAAC CTAGAGGA KATTCACAC 240  
AAAAAGAA TTGGATGCG GGTGAGAGA ACTGCGAG AGCTTAAA GGCAGCAATG 300  
GAAAGATGA TGGAAAAAT ATTAAAAAT GATCAGATGG GAAAGTGT TGTGCTGGA 360  
GGGGTCTG TTGTCTTGG ACGATTGTC TACTATGCT TGGAGCTTC TAAATGAAT 420  
GAGCTATGG AAAAGCTGT AATTGGCTT CAGTATGCA AAGATGAAT TCAATTCAC 480  
TATATTAAT TAGAGAGAG TATTGGTTA AAGACTTTT CTGCAATAGC AATCAGAGA 540  
AGGCTGTTC TCAATGAAT CATGATAGA GGTCTTGGG TGAATATGG TGTACCTTT 600  
GGAGCATGG TTGAGCTGG AATGCTGTA CGATCAATAC CATATGACA GAGCCAGAGC 660  
CGAAAGATC TTGCTTGGT GCTACATCT GGTGTGATGG GTGAGTGT GAGTCTCTG 720  
ACATATTAG GGGTCTCTT TCTATAGA GCTGAGTGT ACGAGCTGG CATGTGGA 780  
GAGCTTACA CTGTGCAAT GTGTGAGCC AGTGAAGAT TTCTGAAGAT GGTGTGAGCC 840  
CTGAGATGG GGTGGTCT GGTCTTGGG TGTCAATGG GATCAATGT TCTTCACT 900

AGCAGCTGG CTGTGCGAC TCTTACTA GTGGCAAT AGGTGAAAT AGTCTTTTC 960  
AAGATTTCC TTCTATATGA TAGCCAGAA GTATCAAGC GTGAGAAAT ATCAGATG 1020  
TAGAGATTC AAAATATGA TCCATTAAC TGAATGCTGA GTATCAAT GATATCAATTA 1080  
AATATTTA TCGAGTTGC AACTGTGTC GAACTGAG GAGAGAGAA GAAATGAAT 1140  
GATCAGCTT CTGCTGTTC TCGATGCA AATATCTGT TTAATGGGC AGATATCAT 1200  
TAATATGTT GTACAGAGC CTTCCTTGA AGTTAGAG ATTAAGAA CA TGTCAATGTA 1260  
TTAAATGTT CGGTAAATGG GATGCTGAG GTCTGCTTT TTTCTGAG AATTAATGCA 1320  
GTATCTCTT CGCAATTAG CAGCAAAAT TTCAATCTC AGTTGAAAT GATTTTAAA 1380  
TGTTTGGT AATGTGAAA CTAAAGTTG TGTGATAGA ATGTAACTT TTTTCTACT 1440  
TTAAATTTA GTAGTTCAG TGAATACTA AATTTTGA AACCTGTGT TGAATTTT 1500  
TTGAGTGC ABAATTTGT AATTAATGC AATAGTAT TTGAGCTTGG GTAAAGGAGC 1560  
CAGAGAGAG GATCAAGTC CAGTCTTGG TTTTATTA TACTAGAC TTAGCACTTG 1620  
TGTATGAT TAGTGAGAG CGATTAAGA AACTGTGCT ATTGGAAC AAGTGTGAT 1680  
TGTAACTTC ATCTGTGGA CTTAAGAAA CTGTCAATC TGAACAGGC AAGGTGATG 1740  
CAATCTGTC CTGTGCTTC TGAATGCT CTTCGATA TGAATGTGT CAATTTGAC 1800  
TTGAGGGA TGTAAATAT ACGAGATC CTGATGAGA TTAATATGT GTGTTTACT 1860  
TTGAAATAT ACGAAAGAA ATACTTTA AACTATCTC AAGAGAAAT ATTCAGACA 1920  
TGAATATGT TGTCTTTC AAGATACAA CAGTATGTC ATGATGCTA AGTGTTTTT 1980  
TATTTTGA TATTAATGA ACTGTCAAT TGAATAGC TTGCTCTT GAGCTTCTA 2040  
AGCTTCAAG CCTTATAGA AAGCTTTT TGTGCTTAC ACTGAAAT ATGAAGAGC 2100  
TTTTCTCT AAGACTTGG GTTCTGGA TTGCTCTCA GACTAGAC TAAAGAGCA 2160  
ACGAAAGAG AACTAGTCT GTCTTAATGA AATATGCA CGGAAATGG TAAATGAGAA 2220  
AATGCTCAT TAAATGCCC TAGAGAACT TTAATCTCT TACACTGTA CAGCTTACT 2280  
TTCTGAGAG AATTGTAAT CCTTGATAC AAGAGTTA TATTTAGAG GACTTAATG 2340  
AAGGG 2345

## (2) INFORMATION FOR SEQ ID NO: 302:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2369 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

60

533

534

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

5 TTTTITTTTT TTTTTCNAG ATCAATGTTT ATTATTAAT TCAGATAAA 60  
AGATAGATA CATATAGGG AATCCCTTAA AATTCACCT TAGAGTTATA CACCATCTAG 120  
TACTTTTGA ATGAATGTTA ACAGACAAA AAAAAATCT TAAACACCTG AAGCCCTAC 180  
TATTACATG GACTATGTTA ATAAAAANT TTGACATTTA ATTGTGTTCA CATATAGAT 240  
TTACATTATG AAACCAATGG TGATCATACA ATAAAGTAT AAAGAAATAG TAAAAATAAA 300  
CTTTAAAAAG CAAAGTTTAA TAGTCTGACA ATCTTAATTA TCTTAATTT ATATAAAAA 360  
TTAAACATA GAGCTTCTG TTACAAANT CTTAATCTC TGGTTGTAA TCATTAATCTG 420  
CTACCAATTT ACATCCACAA TCTCTAGGA CTGACATTTG ATTTTITTC CCAAGATGT 480  
GTGATGAT AAATGACATT TGAAGGACA TATTAAATTA CTTGTGACA GAAAAAGAAA 540  
CTCAAGTTG GTACTGTGTA CAGGCTCTT CCCATAGAA ATTATAAAA CAGTAGATA 600  
AAATTTAAAA AAMTCTTAA AAGGATGTC ATAGCCAAAG AGTACCATA ATGCCACAG 660  
TCAAAAATC CCAGACCAA TGAACACAC ATCTTTTCT TCTCTTCAG GACAGAGG 720  
TGGATTTTC CATCAATTA CAGTATGTA AGCAGAGGAG GGCACACAG TGTACACCTG 780  
ATTAGATCTT GCMAATACT AAGATGGAG CAGGCTGTC CAGAGAGG GATAATTTAT 840  
ATATATTTCA AACTATATAC ACATAAATG GAATCCAGCC CATCCCAAC TGGCTCTGTG 900  
AAACATTTGG ACCTTTTAG TTAAATTTAT ACGATGTGA ATATACAT AGATTATAT 960  
GGGACCAA ATCCAGGGA CATTTATAT TAGTATTTA CTGTCTGTT TCAATTTAA 1020  
AATATTTTT CTAAATATAC ATCTCACTG AAGTCTATGT AAAAAATGTC CTATAGATA 1080  
CAGATATTTA CTTTGTGTA GTTGAGGCC TTTTGTGAC TTCTGTCTGA ACTGTAGGCA 1140  
GAATCTAGA TGTACATGA CATATGGGA AACTCAGCT GAGTCAATCC AAGACTGTG 1200  
GATATAGGA GCTCGAGGT ACTTTGAAG TCAAATGTA CCAGAAACC AANACGTTA 1260  
ACAGTACGA TGGCAGAGG GAATGGAATG CCAATATGG AGTAAACCT TTTTAAAAA 1320  
CAGAAAGG AAGGCTCTC GTACACGAG AATCTGTAG AATCTGAAA AAGAAAGG 1380  
CAGCAACAT TTGTAAAC AGAGCCAT TATAGTGTG GAATGAAA ATTAGGAAA 1440  
ACCAAGATC ACGAGAAA AACTACTGG TTTACTGTAG ABAAGGAGA ATGTTCAAC 1500  
CCAGCAGAG TTACTGTGTG AAGCCGCA CACCGCCA CAGACTCTA TTGGCTTGG 1560  
CTTTGACA TTCACTTCA GGTCTAGT CAGGAATG CCGACTCTC TTGTAGGCA 1620  
AATCACTG CTGTCCAGA AGAGGAAA AAGCATCTC CAGGACTCC GAGGCATAG 1680

CCAGTAAAT GAGGCCAC AGCCCTGT CCATGGGTG AGGTCATTC ACCAATTTGT 1740  
CAAGACGC TTCTGTACT TTCTTGATGA GGGCTGCTT AATGTTGTTA TTGTGAGGG 1800  
GATGTGTT CATGTCAAA AGTAGAAGT TCTGTTCTC TGTGTCAAT ACAAACCTTTT 1860  
CCACAGTT TTAGCTAAT GTTCCGTTA CATTCTTAA CTGTAATGC AATTTAATG 1920  
GATTCATGT CTCACACTA AGTAATCAA TCCAGTCTG GACGTTCT GAGAGCTGAG 1980  
TTTCTTAA ACCTTCAGA GCTTCATCA GAAGAACATC CCGTGTGGA GCATCTGACT 2040  
TAGAATATC CTTCTGTTT ATAGACTTTT TACGTCTCAT TCCACAGCC TCTAGTTGTA 2100  
ACCTTCTCT CAATCTAAT TCAATTAACA TACAGCAAG TAATCCAGAT CATATACAT 2160  
CATTCAAA TGAATGTAA ACCTTCCGG TCTTTAGGC CCAGCAGAG CACTTCTCC 2220  
ATCAGGTCA GCGGTTTC CTTGAGTGG CCTTGTGCT GTGTGCTG CTGTGCGGG 2280  
CGGCTGTGG CTTGTCTCT GCTGTAGCC GCGCGCGCC CCGCGCGCC CTCTGTGTCG 2340  
GCGGTTTC GAGAGGCTC GTGCGCGC 2369

(2) INFORMATION FOR SEQ ID NO: 303:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1181 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

GGAGCTGTG GTTTCAGCTC GTGCGCTCC CCGTGGTTT GCGACGTTTA GCGACTATG 60  
CGCTTGGCC AGCGCGGCTG CGAGACTGGG GCGGTGGTG CTGGTCCGG GTGATCTAG 120  
GCGCTCCCT GGGCTCCAG CTGTTCGGG GTGTAGTGG GAGTCAAGCA CGGTTGCGGG 180  
CCCGAGTGT CCGGAAAGT GCGGCACATG GCGCGCAGG GAGAGCATGG CTCAGCGGAT 240  
GGTCTGGTG GACTGGAGA TCAAGGATT GGCATTTGAG AAGGACAGA TTAATGAGAT 300  
GCGCTGTG ATACTGACT CTGATCTCAA CATTGTGCT GAGGTCTTA ACCTGATAT 360  
AANACAACA GATGATGTC TGCACAGAT GTCAGATGG TGTAGGAGC ATCACGGAA 420  
GTCTGGCTT ACCAGGAG TGAAGGAG TACATTTACA TTGAGCAGG CAGAGTATGA 480  
ATTCTGTCC TTGTAGGAC AGCAGACTCC TCCAGGCTC TGTGCTACTG CAGGAATTC 540  
AGTATGAA GATAGAAAT TTCTTGAAA ATACATGCC CAGTTCATGA AACATCTTCA 600  
TTATAGATA ATTGATGTA GCACTGTGTA AGAAGTGTG AGAGCTGT ATCCAGAGA 660  
ATATGAATTT GCACCAAGA AGGCTGCTTC TCATAGGGA CTTGATGACA TTATGGAAG 720



CACTAAGAG CTTCACTTTT ACCGAAATTA CATTCTAG AAAAATATAG ATTAAATATA 780  
GAGAAAAATT ATTAAAAATG GAGAAATAG GAGAAATAG AGTTATAGC AGTTATAGT 840  
CTGCACATAC ATGTATATCT GAGGCAACT TCTGTGTGTT TTTTCTTC AGCTATAGG 900  
CTTGGGAGG CACTTGGATT AACTGCATC TCCATTTGA TTACTGACG AGACAGACA 960  
CGAAATCTTA TTTTCTCTCT AATATCTCT TCCATTTAG AGACAGACG TCCCTTTGA 1020  
GTACAGATC ATGTGCATC CTGTGTATC ATATGCAATT GCTTTTAACT CATTCTTTT 1080  
GTTAAATTA ATTAATTAAT AATTAAAGCT AGTTCTATTG AATATGAAA AAAAAAAA 1140  
AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA N 1181

20 (2) INFORMATION FOR SEQ ID NO: 304:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1517 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

CTTTTGTGT TCCGACCAAT CCCACTCTC CTCAGCTTC GACCTGACG TTCCGAGAGC 60  
CCAGATCTTG CCCACTCCCG GCGCGGAGCTT AGCCCGGATT TCAAGACAG GAGCCCTCA 120  
GAGACATGCG AACTACAGCG GCGCGGAGCG GCGCGGAGCG AATATGAGC TGGCCCGAA 180  
TGGGAGAGGT TCGAAGAAA CATTCAAGCG GAGGCTGAG CTGTATATTA CATAGAGAAC 240  
ATGATATATA CTCAGAGCTC TTAGCTTGGAG GATTATGGTTC AGCTGCAATCA GCGCCCTGCG 300  
GAGAAAGAG TGAACCTGA TCCAGCTAT GCAACTGCTG CTGAAGAGCA GATATGAGAG 360  
TTCTGTGGCA TGAAGGCTT TAAAGACAG CTGAACCGCG AGTTAGGAGA TCGATATGAG 420  
CAGGCTGAGA AAAAGACAG CTCAGAGCGC TTCAAGTTGT AGCCCAACAT CGACATCTTC 480  
AGACCTTACT TTGATGTGAA GCTGCTCAG GTGAGACAG GCTGCTGCA GTCCATATTC 540  
CTTATCAGAA TGTCTCACTT CCCCCGAAA ATTGCAAGATC AACTATATG AGCTGTGATG 600  
CTGTCTTCA CTCGTGTTC TTATCTTATC CATAGATTA AGAATCTGA CACTATTTATC 660  
CGGAGAGCA CCGTATATG GAGAGCAAT GCAAGCTGCT TCGGCTATG AGTGGAGTC 720  
TCAATCTCA TTATCTTCT TCCATATC TCGAAGCCG AGATACCAAT GCTGAGATG 780  
TTGCATCTGC TGGGCTATAG CCGTTATG CATTGCAATG TCCCTTCAAT CACTATTAAT 840  
ATTCACATCG AGGCGCTCT TACCTCTTC TGGCTTTTGG TGGATGAGCT GTCCACATG 900

CGCATGTAG CAGTTGTGT CTCGTGACG GTGGGCGCCA CAGAGCGCT GCTGCTCTGT 960  
GGAGCCCTGG CTGCGCTACA CATGCTTTC CTGCTTATC TGCATTTTC CTACAGAAA 1020  
GTATATGAG GATATCTGAG CACTGTAGG GAGCCCGACA TCCCGCCAT CGAGAGCTC 1080  
CCGAGAGCA TCCCTGCCAT GCTGCCCTCT GCTGCGCTC CGACAGCTT CTCAGAGCC 1140  
AGAGCAAG CTGTATGCTT GAGCTGAG TCAACTGAC CCGACTGAA ATTCTTGACC 1200  
AGTCTCTTT CCGGAGCTG GAGAGAGAG GAGATCTAT AAAGACATC CTGTATGACA 1260  
TGTTCGTAG ATGGAGTTTG GAGTCCGAC TGAAGCTTAG CTGTATTAAT AGCTCTTGA 1320  
AGCTTGCGC ACTCTGAAA GAGAGAGCG AAGAACTCT GCGAGAGCT GCAAGCTCT 1380  
CGAGCAATG CAGAAATGG GTGAGCTCT TTGAAGAGCC CTGCGACTC AGCCCTTCT 1440  
TCTCTTAA CTGTCGACA TTGCTTCTCT AATATATAC TTGTAATTA AAAAAAAA 1500  
AAAAAAAA AAAAAAAA AAAAAAGG GAGCCCG 1537

20 (2) INFORMATION FOR SEQ ID NO: 305:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1493 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

TGCATGCCA AACCAATGC TCCCAACA AATCTTACAG ATCCCAATAT AATATTTAG 60  
TTATATTTCT ATTCAATCA TTATGTAAA TACCGACTC AGTGCATGC TTAAATATG 120  
TTTATTTCCC TTACTTATC TTGCTTAT TTTTATTTTG AATATGAGT GAGCAAAATA 180  
ACATATTC GCTGTAAGCA ATTTTGTGA CATTCTTGT TACAAAAA TCTTATATCA 240  
GATATATCT GAGTGTACA AACAGCTCT AATATTAAG AGATGAAAC TCTTACAGA 300  
GCTGAAAT TTAAAAAGAG AGAGAGAGC CAAAGAGATG CATGAAAGCA AAGGAGAGG 360  
GAGAGAGAG ATTGAGCTC AAGAAAAAG CAAGCGGAA AGAGATGCC AGAAAACTT 420  
TGAAGAAAT CGAGATGCT GTGTGACAG CTGGGAAAC TTCCAGCTA ATACAGAGG 480  
GAGAGAGAG AAGAAAAAT GAGCTTCTT GAGAGCAAG AAATGTAATA TCGAGAGAG 540  
TGAATGAGG CCGAATCTA CAGAGAGAG AGCTTTCCC TGTATTTCC CTTCCTGCTT 600  
CGAGAGCTC ATTTCTTCT CCGACTTCA CCGCAACATA GATATGAT TGTCTTTTGA 660  
TGCATTTCT TTTCATAGG ATTTAAATC GATCAGAGTA ATCTTTTGT AGATGTAAT 720  
GAGGCGCTTG GTTAAAAA AGACCTTTC CTCCTCTGC CCGTACACA ACCATATTA 780

5 GAGGTGCCA CCATTGGTGC TGGCTCTCT TCCACAGGC TGTACTCAG TTTTITGAC 840  
TTCACTGAAT TGTGATGGTT AGAAACTTGG TGGATAGTTT GTGGAAATCA TCCATATAA 900  
CATACTGCTT AAACAGTGT TGTGTGACT TCAGAGCAA GCTTGGAAGG GGCACTTAG 960  
GAGCCCTT GCTTCAGTT GCTGCTCTCT GGTGTGCTC CTTTGGAGG CCCAGATAG 1020  
10 ACAGGGAACA CTTGTGAGCA CACAGAGCAG CATCTGATGC CTTGTGTGT TTTGCAATG 1080  
CCCCCTGCT ACTGACCAAT CAGTGTGCA TGAAGGCCAC GCCACCCAA CTTTTCAGTT 1140  
TCCAAAGAGC TAGCCCTGCT CACCCAGTA CCATGCTCTA GCTGTCTGCG ATTTGTAGT 1200  
GTAATATTC TTTATGTATA ATAAATTTT ATACCCAGC CATTCATGTA CTTTCTCTG 1260  
TACTCTCCT TGTGGTGGC TTGTCTGCT TGGCTGAAC CCMAATGCT TTGGGTTGG 1320  
ACAGAGCTGG CTGAACCTTA GTTCTCTCAT CTATGAAAG GGAATGCA TTACTGACC 1380  
AGCTTTTAGG GCAGATTTGC CATGCAATAT ACAAGTAAAC TACCATAGT CTCCTGGGT 1440  
25 ATTGCAATA TCCATATAT TCTGTGAAA ATGAAGTAC TGAATGTTT GAG 1493

(2) INFORMATION FOR SEQ ID NO: 306:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 577 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

40 AATTGGCAG AGCAATATTA TACACTATAC TGGCAATTAC TGTTCACCC AGCCCGAAA 60  
GTCAAGATG TATATTGGA AATTACAC TCATCTACA TTGGTTCCA GHAACCTTC 120  
ATAGCAATT ACCCAAGAT CTACAGGAT GATAGAACA CTTATATG TTTATGACTT 180  
CACTATATCT TATATTTTA TTGTTATTT TGTGTTTAT GCACAGTAC TTCAACCTT 240  
AAACTGCTT TGAATTGGT ATGTAACTT TTAAACATG CAGATCAGT TAGAACTGT 300  
CATAGAGGA GAGCTAGAA TCCAGTACA TGAATTTTAA ATAACTGTC TTGTTTTG 360  
ATGTTAACA GTAAATGCC ATAGTAGCA AGAACACAGT GATTATATC ACTATAGT 420  
AGGAATTTCA TTTTAAATC ATCTTATGA AGATTAGAA CTCATTCCT GTGTTAAG 480  
55 GGAATGTTA ATTCGAAT AAACATTTCT GAACAAATG YTAATAAAA AAAAAAAAAA 540  
AAAAAAAAA AAAAAAAAAA AACTCGA 577

## (2) INFORMATION FOR SEQ ID NO: 307:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2860 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

5 GTGTGAGCC GTCTGCAAT ATGCTCCCC CGGCTGCCA GRMRATCTGT CMCRTGTCG 60  
TAGCCTCTCC TGAACAGGGA GAGTTAAGCT CCGTTCCTCC ACCGTGCGCG CTGGCAGGT 120  
GGGTGAGGG TGAACAGGAG ACCAGAACCT GCTTGTCTGA CTTTATGCT CAGAGCTGG 180  
GAGGAGGTT CGGCGCTCC TCTGCTGCA GCGCGGCGAG CCGCTGCGCG CTTCACTTC 240  
TCCCGAGCC CCGTCTACTG AGAAGCTGCG GATTCGACG AGCTGCGCG CCGTGGCTTC 300  
AGCTGCGGG GCTTCCATGC AGGCCAGAC CGAGCGGAC TGGGAGGAA GACAGGACC 360  
TTGACATTC CATCTGACA GAGTCTCTGG CTGGAACCGA CAGCTCTCT CTTCTAGGA 420  
TGACCTACC CTTCACTCT CCAATTTTCA GGTGTGAGAC ATTAGATGGA GGGCAAGAG 480  
ATGCTCTGA GCGGAGAGA GGAAGCTGG ATTTTGGAG GGGCTGCTCT CCGTGGAGT 540  
60 CACAGTTCCA GGGCAGGAG CGGAATTCG CCGTTCAGA TAAGATCAA CTTCCACTA 660  
CCGAAGGGA ACAGTGTCCA GTACGCGGA TCCAAACGA TTGACCGAG ATCGCTCTT 720  
CAATGGGTC TCCCGGGTG TCCCGAGGA TCTGCTGGA CTTTCAAGT ACCTGAGCA 780  
GACGAGCAG TACTCTACG ACTTGGAAA TACACAGAG GCTCCACAG TAAGACGGC 840  
TGATGAAGC TGTCTGAAA CTTTAAAGAC GGGTCAATG CTTGCAATCT GCACTGCTG 900  
CAGATGACC GGGACTCTGG CAATCTCTAG CCGTGTGAA ATGCCAGTG CACATGAC 960  
TATTAACCG GCGACAGCG TCTCACATC GCATTCAGA AAGAGAGTC TGCAGTGTG 1020  
45 GAGCTCTG GTGAGATG GGGCAATGT GCATGCGCG GTCTGCGGG ACTTCTTCA 1080  
GAGGGGCAA GGGACTTGT TTTATTTGG TGAAGTACC CTCTCTTGG CCGTGTGAC 1140  
CAGAGTGG GATGTGTAA GCTACTGCT GGAAGACCA CACGAGCGG CCGAGTCCA 1200  
GGCACTGAC TCCAGGGA ACACAGTCT GCATGCGCTA GTGATGATC TGGCAACT 1260  
CAGCTGAGA CATTGCACT GTGACAGCA TGTATGATG GCTCTTCAA CTKGSGSC 1320  
55 SCCTTCCG CTACCTGCA GCTTGAAGAC ATCCGACCC TGGAGATCT CAGCGCTCTG 1380  
AAGTGGCG CCAAGGAGG CAAGATGAG ATTTTCAGC ACATCTGCA CGGGAGTTT 1440  
TGAAGACTA GCACTTTC CCGAAGTTC ACCGAGTGT GCTATGGGG TGTCCGGTG 1500  
TGGCTGTAT ACTTGGCTT TGTGACAGC TGTGAGAGA ACTCAGTCT GGAATCATT

1560 GCGTTCATAT GGAAGAGCCG GAGAGAGAG GAAATGATG TTATGAGAGC CCGTAGAAA  
 1620 CTGTGTGAGG GAAATAGGGA TTGTGTATC CCCAGATTC TTATTAACCT CCGTGTAAAT  
 1680 CTGATCTACA TTGTATATCT CAGCGATCT GCGTCAATC AGCTTAACCT GAAAGAGAG  
 1740 GCGGCGCCCTC ACTTGAAGAG GAGATTTGGA AGCTGAATC TTGTAGAGGG CCAAGATCTT  
 1800 ATCCGTGAG GGGGGAATTA CCGTCATGAG GGGGCAAGTG TTGATCTCT GCGGCGCCCA  
 1860 CCGTGTATAT TTGATCTCT TTATAGAGAG CTATCTTGA AATCTCTTC CTGTTCAGAG  
 1920 CCGTCTCTCA GATGTGTATC CAGATGTCTG TTATTTCTCTG GCGATGAGT GGTATCTGAC  
 1980 CCGTCTCTCTG TTGTGCGATG TTGATGAGCT GCGTGAACCT GCTTTACTTA TTACAGCTGAC  
 2040 GTTCCAGAGC AGAGGAGATC TTACAGTTCA TTATCTCTGA AGCGTGTATG AGCTTGAAGC  
 2100 AGAGAGCTTG GCGGCGCCGAA GCTTCTTACAG GCGCGAATC GAGAGATCA GTTACAGCCA  
 2160 TTGAGAGGACA GAGAGAGAG GCGTACAGGG CCGATGACAG GGTATATCTG GAGAGCTCTT  
 2220 TTGAGACTTT CAATTTACG ATTCGCAATG GCGAGCTGAC CTTCAGAGAG CAGCTGACCT  
 2280 TTCCGCGCAT GTTGTGTCTG CTGTCTCTG GCTAGATGCT GGTATCTTAC ATTCCTGTAC  
 2340 TTGAAATATCT CATTCGCTCT ATGAGAGGAA GGTATAGATG TTGCGATCTGA CAGCTGAGAC  
 2400 ATCTGAGAG TTGAGAAAG CATCTCTGTC CTGAGATAG AGAATGACTA TTGTGTATGAC  
 2460 AGGAAAGAGC AGCGGAGAG TTGTATGCTG AGCGTTGGA CTAAAGCCAG ATTCAGAGCC  
 2520 CAGTGAAGCC TTGTCTCTCA GGTATGAGGA GGTATAGTG GCTTCATATG GAGAGAGCC  
 2580 TTGCTTAACCT GTTGTAGAGC CCGTCAAGGG CAGATGTCCC TTGAACTTTC GAGAACTCTG  
 2640 TTCTGACTTC CCGTCCAGAG GAGATGAGG ATGTGTCTTC TTGAGAAAGC TTATGTGCGG  
 2700 TTGAGACTCT CAGATCTGAG TTATGAGCCA GATTCAGAG GAGCGCGAG GAGAGAGAG  
 2760 AGATATCTTC CAGAGATC TTGTGACTCT GGGTCTCCAG TTAAATCTG TTGCAAAATAT  
 2820 AATTTTTCAC TTACTGAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGAG GGGGCGCGAT  
 2860 AAGCAATATC GCGCTATAG TTAGTCCMA TTAGCAATAA

50 (2) INFORMATION FOR SEQ ID NO: 308:

55 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 876 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

60 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

60 CTGTCTGTCT CTGCGCTGCT GCTGAGCTG CTGAGCTCTC TTGATCTATC AGCTGTCTTC  
 120 GAGAGAGAG GATCTGAGAG TTGTATGATC CAAGAGCCCT GGTAGAGCTG AGCGAGAGAT  
 180 TTGTGCGCCCG GAGATCTCTA CAGCGCCGAA TTGCGAGAG TTAGTGAAG CAGATGAGAG  
 240 AGAGAGAGAG GAGGAGAGAG GCGGCGAGAT CAGGCGATATC CTGGAACCTT CCGTTGAGACT  
 300 CTTCAAATATC AGCATGCGGA TTGGGAGAGCT GCGCTTCAG GAGAGCTGAC ACTTCCGCGG  
 360 CATGTGTCTG CTGTGCTATC TTGCGTATCT GCTGTCTGAC TTACATCTCTG TTGTCAGCAT  
 420 GGTATGAGCC CTCAATGAGC GAGAGCGACA AGAGATGAGC CAGTGAAGC TTGAGCATCT  
 480 GGAAGCTGCA GAAAGAGATC TTGTGTCTGAG AGATGAGAAA TTGCTATTTAG TTGTGCAAGGA  
 540 AGAGAGAGAG GCGAGATGAT ATGCTGAGAG TTGTGACTTA GCGAGATGAC AGCGCGAGAT  
 600 AGCGATGATG CTTCAGAGATG GAGAGAGTGA ACTGTGAGCTT ATTCAGAGAG AGCGTCTCTA  
 660 CCGTGTATTA GGAAGCGTCA GGGGCAAGTG TTGCTGAGAG TTGTGAGAGC CCGTGTCTGAG  
 720 CTTCGCGCTCC CAGAGAGAT GAGATATGAT CCGTGTAGGA AATCTATATG CCGTCTGAGC  
 780 TTCTGTGATC CAGCTATAG CCGAGATGGA GCGAGAGCC AGAGAGAGGA GAGAGAGATC  
 840 TTTCAGAGCA CATCTCTGAG CTGTGAGGAT CCGATGAAAT CTGTGTGCGAA AATATATATT  
 876 TTACTAATAT AAAAAAAAAA AAAAAAAAAA ACTGGA

30 (2) INFORMATION FOR SEQ ID NO: 309:

35 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2025 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

60 CATGAGCCCG CTGATGCAAT CCGGAGAGC CTCTGATTC AGCGTACCTT CTGTGATAG  
 120 CAGCGAGAGC GCGTCCGACA CAGAGAGAGG CAGCGAGAGC GCGTCCGACA CAGAGAGAGG  
 180 CAGCGAGAGC GCGTGTGACA CAGAGAGAGG GCGGCGAGCT GAGCTGAGCC CCAAGCTGAG  
 240 TTGTCTGAG AGAGAGCGCG GCGCGAATTC ATGAGATCTT CCGTGTAGAG GCGCTGCGCA  
 300 GCTGCGCCCG CCGAGATCTG ATCTCTGAG TTGCGCGCTC CTCCCGGAGC CCGTTTGTGCG  
 360 CCGTGTGCTC CATATCTGCG CTAACTGATG AATTAATCAA AGCTTATTTT GTTAAAGATGA  
 420 GCTCTGTGAG AGAGAAATGA GATCTATGAG GTTGTGTCCC TTGTCAAGAG CCGGATGTGCG  
 480 GTTGAAGCAA GAGAGAGAG AGCGATCTCC GATGTGATC CTCTTCAATT AACGAGTGGC  
 540 GGTGTGAGAG TTGCTGTCCC TTGCTGTAGAG AGAGCAACT GCGAAAAACA AGAGCGTATC

AGCAGACACT TCACAAAGCC AAGCCTAGGC GCGCCTGAGC ATCTGTGTTT AAGCGGTGC 600  
CTGTCTAGAA GCGCAGCGCC CGACTTCCCG TTCTCTCTTT ACTGTAGGAG AGCTGTATCC 660  
AGTTTCCGGA AACAAATCC TTTTCTCATT TCGGAGGGG GGTATATAGT ACATCAGGC 720  
ACCTCTTTTA AICAGGCAAA ACAGAGAGGG GGAAGAGGTG GGAATTCATCT CGAGCCTAGA 780  
GGCATTTTGA ACACAAATC FACGTAGTTA ACTTGAAGAA ACCGATTTT AAGTTGCTG 840  
CATCTAGAAA GCTTTGATG CAGACAAA CAAGCTTGAT TTTTCTAGCA TCCCTTAT 900  
GTCCAGCAAA AGCAGGCRAC AAAATCTCTT GCTTTTACAG ACAAATATAT TTCCAGCAAC 960  
GTTCGGCATC ATGTTTTTGG AAGGCTTAG TTCTGCTTTC TCCCTCTCTT CCACAGCCCC 1020  
AAGCTCCAC CCGTCATAGA TCAGCAGTG ATATATCTTG TTCCAGGGAGA AGATCATTTA 1080  
GATTTGTATT GCATCTCTTA GATGAGGG CAACATTTCA CAGCTGCTCT GCTGTGATG 1140  
AGTGTCTTGG CAGGGGCGGG AGTAGAGCA CTGGGGTGGG GCGCGAATGG GGGTTACTGG 1200  
ATGTAGGGA TTCTTGTGTT TGTGTGTCAG ATCCAGTSCA GTTGTGATTT CTGTGATCC 1260  
CAGCTTGGTT CAGGAAATTT TGTGTGATTT GCTTAAATCC AGTTTTCAT CTTCGACAGC 1320  
TGGGCTGAAA CGTGAATCTA GTAGCTGAAC CTGTCTGACC CGGTCAAGTT CTTCGATCT 1380  
CAGAACTCTT TGTCTTGTG GGGTGGGG TGGGAATCTA GGTGGGAGC GGTGGCTGAG 1440  
AAATGTAG GATTCGTGAA TACATATTC ATGGGACTTT CTTTCCCTCT CTTGCTTCT 1500  
CTTTTCTGC TCCCTTACT TTGCGCGAAT GGGGAGCAGC CACTGAGTT TCTGGGCGGC 1560  
CAGTGGGCT GCGAGTTCC TGTACTACTG CTTGTACTTT TTGATTTGG CTCACCTGG 1620  
ATTTTCTCAT AGGAAGTTTG GTCAGATGA ATTGATATTT GTATGTACG CACTGGGACC 1680  
CGAGGATTTT TGGAGCCCGG CAGTTGGGAG GAGGAAGTAG TCCAGCCTTC CAGGTGGGCT 1740  
GAGAGGGAAT GACTGTTTAC CTCGCGCA TCACCTTGGG GCGCTTCCCT GCGCTTGGT 1800  
AGAAAAGTGG GGAATCGGG CAGAGAGGC TGAGTAGGGA TGGGAATCTA TTGTGCACAA 1860  
GTCTTTCCAG AGGAGTTTCT TAATGAGATA TTTGTATTTA TTTCGAGCC AATAAATTTG 1920  
TACTCTTCCA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAACTC 1980  
GAGGGGGGCC GATACCANT TCGCGTATA TGATCTTAAA CAACTC 2025

(2) INFORMATION FOR SEQ ID NO: 310:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3026 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi.) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

5 TAGGCAGAC TGAATATATCC TAACCCCTTA AGCTCCAGGT GCGCTGTGCG ACAGCAACT 60  
GGACTATAGC AGGCTTGGC TGTGTCTTC TGTGTATAGG CTCACCTTTT CCCCCAATTC 120  
TTCTCTGGA GCTTTCGAGC CAGGTGCTA AAGGATAG GTAGGAGCC TCTTCTATCT 180  
AATCTTAAA AGCAATATGT TGAACATTTCA TTCAACAGCT GATGCCCTAT AAGCCCTGCC 240  
TGGATTTCTT CTTATTAGCC TATPAAGAGT AGCAAGATCT TTACATATTT CAGAGTGGTT 300  
TCAATGCTCT CTTACCTCTCT CTATGCGCC CTCATTTTAT TTGACTAAG CATCACAGC 360  
TGGCACTAGC ATTATACCA GAGTATGAGA AATACAGTGC TTTATGGCTC TAACATTAAT 420  
GCTTTCAGTA TCAGGCTGC CTGAGAAAG GATGCGAGCC TCAGGCTTTC CTTATGTCTT 480  
CAACACAG AGCTCTTGA TGAAGTCAAT CTTTTCCTCC TATCTGTGTC TTCCCTCTCC 540  
GCTCTTAT GGTAGTGGG TACCCAGGCT GGTTCCTGGG CTAGGTAGTG GGGACCAAGT 600  
TCAATTAATC CTTATCAATT CTAGCATAGT AAATCAAGT ACCAGTGTTA GTGGAGAGAG 660  
CTGGGTTTTC CTAGTATACC CACTGCATCC TACTCTTACC TGTCTCAACC GCTGCTTCCA 720  
GGTATGAGC CTGCTAGTG TGGATTACC TGATAGGGA GAGGAAATA CAAGAGGCG 780  
CTCTGTGTT CCGGCTCA GCGAGCTGCC CACAGCCCAT AAACAAATTA AACAAAGATA 840  
CTGATCAT TTTTATCTG GGTCTCTTC ATTCCTACTG CACTTGTGTC TGTCTTGGCT 900  
GACTGGGAC ACCCCTTAC TACAGATCT CACAGGAGA GTGGAGACTG TCCACTTCTA 960  
GCTCGAAT TACTGTGTA ATAACTTTC AGAATCTCTA CCATGAGTG AAAATGCCAC 1020  
ATTTTCTTT ATATTTCTA CCGATGTGG GAAAACTGG CTTTTCCTCA GCGCTTCCA 1080  
GGGCTTAAA CTCAGCCCT TCGTATGCAA GTCCATCAG CTTATTTT TTTTAAAGAA 1140  
AATTTGCACT TGTTTTCTT TTTACAGTTA CTCTCTCTT GCGCAAAAT TATTAATCTT 1200  
AAGTGTAAA AAAATGCTTA ACAACAGTT CTGTCTGTGA AAAATATGTA TTATACATCT 1260  
GTATTTTAAA ATTCTGCTC TGAATAATGA CTGTCCCAT TCTCCACTGAC TGCATTTGGG 1320  
GCGTTTCCA TTGCTCTGA TGTCTTTTAT CATTCAGGC CAGTGCAGAG AGGGAAGAGG 1380  
GAGACAGGG GTGCCAACA CTGTGTGTC TTCTGACTG ATCTGAAACA AGAAGAGATA 1440  
ACACTGAGGC GCTGCTGCC ATGCAGACT CTCGAAACA CTTATCTCC TCGAAGAGTG 1500  
GGCTTTCAG GGTCTTACT GGGAGCAGT TAAGCCCTT CCGCAACCT TCTTTTTTC 1560  
TTCTTACT CTTTGGCTT CAAAGATTT TGGAAAGAA ACAATATGCT TTACACTCAT 1620  
TTTCAATTC TAAATTTGA GGGATACTG AAAATPACGG CAGGTGGCT AAGGCTCTG 1680

	TTAATGTTAG GGGAGAGGAA ACTTTAAAGT TGAAGATTA AAAAGAGATC CCGTAACAA	1740
	AAAAGAAAT AAGACTGTC TTCCATTTTG CCACTTTCC TGTTCATAC AGCTACTAC	1800
5	CTGGAGACAG TTAATTTTCA TTACCAAG AAAGTGGATC AACTAGCTC TGAAGACTG	1860
	AGTACTAGG CCACTCAAT CACCTTACA GATGCCAGG AGTCCCAAG AATGCCACT	1920
	CGTTAACTG AGCTAGAAA ATTAACCTGG GCAAGTAGG CAGAGAAAT GAGAGAAAT	1980
10	CAATCTGGA GGTCAATGAC AAGATTAATA GACAAAGAG GAAAAAGTA TCAAGAGAG	2040
	AAAAGAGATC ATTATTTTGG GTCTGAAAG AAAAGCTTT GCTATCCAGC ATTTACTCT	2100
	AGTACCTGTA AGCATTTTGG GTCCAGAAAT GGAATAAAA ATCACTATTI GGTAAATATA	2160
15	TAATCTCCT TCCCTGAGT CAGTTTTTT AAAAAGTTAA CTCTTAGTTT TTACTTTGTT	2220
	AATCTTAAA GAGAGGAGG CTGAGCCAT TCCCTTAGG AGTAAGATA AAAAGTAGG	2280
20	AAAAAGTTCA AAGCTTAAT AAGATCAGG CTTCACAGG TATTAAGCT AAAATTAGA	2340
	AGTCAATTA GCAAGCTGG AAAATGATCT AGTTCTGAT AGCTACCCAG AAGAGAGTG	2400
25	ATTATTAAT TTGAATTCA AACTACTTTC TTATATATAC TTGGTCTCC ATTTTCCCA	2460
	GAGAGAGAA TATGTCCCC CCAACTTTC TTGCTTCAA AATTAATC CAGATCCCA	2520
30	AGATCACTCT ACAAATTAAT TTGCAAGAC ATCTCTTAC CCGATGCTT GTCTGAGCT	2580
	CACCCAGCT CAACTCAACA ACTGTGTTT GAAACCACT GCTTAAGCT TGTGGGAGG	2640
	GGGATTTAG TGAATAGGA GACCCAGAG TGAATAGAA AAGGTAGAGA CTTCACATG	2700
35	TTGGCTCTC AGACTTGAT TGAAGCCA GACAGTCCA GCAAGAGAG ACTTGCCCA	2760
	GGAAAACT GTGGTTTGG CTATTTTGG TCCAGAAAT AGGTGGACA GAGCTGTG	2820
40	GGTGCATGG AAGATTTGG AACTGTTAT GTTGTTATTC TCCAGCTG AATTTTGGT	2880
	ATGTAAACA GATTTTCTG TAACTTAAT GTCTGTATA ATTAAGAGG TTAACTAGT	2940
	AAAAATTTCA ATTAAGATC AAAAAAAA AAAAAACT GAGGGGGG CCGGTATCC	3000
45	AATTTTCAA ATTAAGATG TATTAC	3026

50 (2) INFORMATION FOR SEQ ID NO: 311:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 712 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

60 GAGAGCTTG TGTCACTTA CAACTGGGT GAGAGGGTG CCAAGACTT GTTTCCTTT

	CTCTCTGGG AACAGAGGCT TTGTCTCCC GAGTTGGAGC TGTGAATGG TGTGGTCTCT	120
	GTGTGTCTCT CCAATCTGAG CTCTCTCTG GGTGGAGCT TGTGGACCA GCACTGAAA	180
5	CTGTGCTCTC TGTGAAGTC GGTGTGACC TTCCACTGG GGGGCTTAGC CTGTAGACT	240
	GCTTGTGCT TCACTGTGGA CACCTGAGG GCGAGATGG AACTGTGAGC AATTTGAGA	300
10	GGTGGAGCT TGTGAAGCT ATGTATGAG CACTTTGG GAGGCTGAT GACAGAGTC	360
	AACTTACTG GGTATATGG CTGGAGCAG CTGGCCCCCA GGGCTTGAG GCGAGACT	420
	AAAGCTTCT GGCAGAGCT GAGCTGCTGG GAAAGCTGCT GTTGGAGCT CTGGGAGGC	480
15	CTGTGTATGG GGTGGGGGCC AATATCTGAG TTCTTGTCC TGTCAATCT CTGTGCTTT	540
	CCCTGTCTGT AACTGAGCT AACAGCCAGC AACTTCTCT GAGCTGAGTG GGTGAGTGG	600
20	TCAATTAAGC CAGATGTGCC TGTGGCCCA AAAAAAAA AAAAAAAA AAAAAAAA	660
	AAGTGAAGG GGGGCCCCGT ACCGAATGG CCGATATAGA TCGTAACAA TC	712

25 (2) INFORMATION FOR SEQ ID NO: 312:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1289 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

	CAAAATTTCA GAATTTTCA GAGCCAGAA GAATATCAA CAATATTTT TGAAGTAT	60
	TTGCCAACT TCTGGTTGAG CTGCAAGAA AATTTATGG TGAAGCTTT TGTGTTCCC	120
40	GTATTTGCT TTTTGGTTGG TTTTGTTTTCA TGTGTTGCTC TGTAAAAATA	180
	TGCACTGAA CTACATTCAG AAGAAATAT TGTCTACATA GAATATATA TGAAGTTGCT	240
45	ACATTAATCT GATGAGGAAA AAAAATCTTT GCAATCTTT AAGCAATAT GTTGTTTTTC	300
	TGTGTGTTT TCCGTGAGG AAAATATCAG TATTAAAGG AACAAATAT ATTCAGATGT	360
50	TGAAGTAT TAAATGTTTC TTGCTGTGTA TTATACATA TGTGATTTT GCAAAATAT	420
	GCTTTTTHA AAGAAAGCTA TAAATGATA CATATGAAA AAGAGATGG TGAACCTTT	480
	GTGCTCTTC CACTGAGAGT AAGAAAGAGC ATGTATATCC ATTTCTTTC AACTTCTTCT	540
55	TCTTATCTTG TTAATACGCT TTATTAATTT TTGAGAGGG ACAAAGATGG GCGAAGGGA	600
	AAGAGAGCT TATTGACTA ACAAAGCCCT GTTGTATCCA CCAAGCTTT GCGTTGGTGG	660
60	GAGAGCTTC AATGACAGG GCGCCAGAG GAGAGAGAA GTGGGGGAAA GCGTGGGCTC	720

5  
10  
15  
20

CGCGCTGGGG AGCTTTTGCCA TGTGAGCCAC GCGCTCTCCA GCGCATGCTC CTTCAGACTTG 780  
GAATGTGCA CCGAGGCCCT TACACGAGCC CTCAGGATC TATAGACTT GATCTACTC 840  
TAAAGGATA TTATATCCAA CCTCACTACA TTGTAGTCA GTCCAAGCAG TAACCTGAA 900  
ATGGGGGTGT TCCGACTTTC AGCAGATGG CCAGCGGTC CCGTGGGGGC TGTGGAGCG 960  
GGCTTATCCT TGTCTGTGCG CAACCTGCCC GTCCGACTTC CTCGCCCCCC ATGGGTTGAC 1020  
CCCGTCCCTG TGTGTGTG TGCTTAGTGG TGAGTTCNCG TAAAGAGCA AATCAGACC 1080  
CGTGGGCCCA GCTGGGAAGG TCGCTGGACA AGGCTCCGAC GTCTCCGAG TCCAGCCCTT 1140  
GGATGGGCAAT TCGTTTGTGT GCGTTATGCC TGGAGATCT GTATACGGCT CGCTTATAGA 1200  
AATATAGCCT CTTCATGCTG TATTAAAGG ACTTTTAAA GCAGAAAAAA AAAAAAAAAA 1260  
CTTGAGGGGG GAGCCGTAC CCAATTTC 1289

## (2) INFORMATION FOR SEQ ID NO: 313:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

Met Phe Leu Ile Phe Val Tyr Phe Leu Lys Ile Leu Phe Ser Ser Ser 15  
1 5 10

35  
Leu Pro Phe Leu Trp Leu 20

## (2) INFORMATION FOR SEQ ID NO: 314:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 128 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

50  
Met Met Phe Leu Thr Gln Gly Gly Pro Leu Pro Ser Thr Arg Ala Arg 15  
1 5 10  
Pro Thr Cys Gln Ala Gly Ala Leu Pro Lys Pro Ser Gly Leu Leu Gly 30  
20 25

55  
Val Thr Cys Trp Asn Gly Leu Lys Gly Pro Leu Cys Gly Asn Arg Cys 45  
35 40

Ser Pro Asn Thr Leu Leu Ala Ala Arg Gln Ala Leu Trp Lys Gly 60  
50 55

60  
Arg Gly Arg Thr His Gln Asp Leu Pro Gly Pro Leu Gln Gly Arg Gln

65  
70  
75  
80

Leu Gly Pro Gln Pro Lys His Leu Leu Leu Leu Pro Pro Arg Gly Gln 95  
85  
5  
Glu Ala Ser Trp Ala Ser Ser Leu Pro Gly Gln Gly Pro Leu Pro Leu 110  
100 105  
10  
Pro His Ile Asn Cys Thr Val Phe Ser Leu Lys Ala Ser Phe Ile Lys 125  
115 120

## (2) INFORMATION FOR SEQ ID NO: 315:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

25  
Met Gln Phe Leu Leu Thr Ala Phe Leu Leu Val Pro Leu Leu Ala Leu 15  
1 5 10

30  
Cys Asp Val Pro Ile Ser Leu Gly Phe Ser Pro Ser 25  
20

## (2) INFORMATION FOR SEQ ID NO: 316:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

40  
Met Asp Gly Phe Ser Ser Arg Leu Phe Ser Ser Leu Pro Phe Val Ala 15  
1 5 10

45  
Leu Gln Trp Phe Ile Val Ile Ser His Leu Leu Ser Leu Ser Leu Ser 30  
20 25

Ala Cys Cys Tyr Gln Thr His Cys Ser Leu Xaa Gln Leu Ser Ser Ala 45  
35 40

50  
Phe Ser Xaa Met Gly Glu Ser Cys Val Gly Glu Arg Glu Tyr Xaa Phe 60  
50 55

## (2) INFORMATION FOR SEQ ID NO: 317:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

5 Met Pro Leu Ile Asn Leu Leu Leu Tyr Tyr Val Pro Asn Gly Gly  
1 5 10 15

10 Lys Gln Asp Lys Lys  
20

(2) INFORMATION FOR SEQ ID NO: 318:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

15 Met Gly Arg His Leu Val Leu Val Met Phe Ile Thr Thr Ser Leu His  
1 5 10 15

20 Ser Gly Thr Pro Val Pro Glu Asn Val Ile Cys Gly Val Thr Lys Gly  
20 25 30

Pro Gln Gly Lys Lys Lys Lys  
35

(2) INFORMATION FOR SEQ ID NO: 319:

35 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

40 Met Leu Trp Trp Ser Arg Asp Tyr Thr Met Val Phe Leu Leu Phe Thr  
1 5 10 15

45 Met Val Phe Thr Gly Asp Leu Val Ile Arg Gly Arg Thr Glu Leu Ser  
20 25 30

Leu

50 (2) INFORMATION FOR SEQ ID NO: 320:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 88 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

60 Met Val Cys Ser Ser Leu Cys Asp Ile Gly Gly Ile Ile Thr Pro Phe

1 5 10 15

Ile Val Phe Arg Leu Arg Glu Val Trp Gln Ala Leu Pro Leu Ile Leu  
20 25 30

5 Phe Ala Val Leu Gly Leu Leu Ala Ala Gly Val Thr Leu Leu Leu Pro  
35 40 45

10 Glu Thr Lys Gly Val Ala Leu Pro Glu Thr Met Lys Asp Ala Glu Asn  
50 55 60

Leu Gly Arg Lys Ala Lys Pro Lys Glu Asn Thr Ile Tyr Leu Lys Val  
65 70 75 80

15 Gln Thr Ser Glu Pro Ser Gly Thr  
85

20 (2) INFORMATION FOR SEQ ID NO: 321:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

25 Met Gln Pro Gly Ala Gly Val Leu Val Leu Gly Leu Leu Pro Pro  
1 5 10 15

30 Pro Gln Ser Pro Ser Leu Ser  
20

35 (2) INFORMATION FOR SEQ ID NO: 322:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

45 Met Thr Phe Thr Leu Gly Asp Ser Gln Val Leu Leu Ile Asn Leu Phe  
1 5 10 15

Pro Ser Met Pro Ser Gly Ser Cys Ala Arg Pro  
20 25

50 (2) INFORMATION FOR SEQ ID NO: 323:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

60 Met Cys Leu Glu Cys Trp Ala Glu Asn Leu Leu Gly Pro His His Thr Ser

549

1 5 10 15  
 Ser Leu Leu Asn Pro Arg His Leu Pro Ser Ile Pro Ala Met Phe Pro 20 25 30  
 5 Val Ser Ser Gly Cys Phe Gln Glu Gln Gln Glu Met Asn Lys Ser Leu 35 40 45  
 Val Ser Cys Leu Phe Val Leu His Phe Val Leu His Cys Ile Phe Xaa 50 55 60  
 10  
 15  
 (2) INFORMATION FOR SEQ ID NO: 324:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 196 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:  
 20 Met Leu Ser Thr Ser Glu Tyr Ser Gln Ser Pro Lys Met Glu Ser Leu 1 5 10 15  
 Ser Ser His Arg Ile Asp Glu Asp Gly Glu Asn Thr Gln Ile Glu Asp 20 25 30  
 30 Thr Glu Pro Met Ser Pro Val Leu Asn Ser Lys Phe Val Pro Ala Glu 35 40 45  
 Asn Asp Ser Ile Leu Met Asn Pro Ala Gln Asp Gly Glu Val Gln Leu 50 55 60  
 Ser Gln Asn Asp Asp Lys Thr Lys Gly Asp Asp Thr Asp Thr Arg Asp 65 70 75 80  
 40 Asp Ile Ser Ile Leu Ala Thr Gly Cys Lys Gly Arg Glu Glu Thr Val 85 90 95  
 Ala Glu Glu Val Cys Ile Asp Leu Thr Cys Asp Ser Gly Ser Gln Ala 100 105 110  
 45 Val Pro Ser Pro Ala Thr Arg Ser Glu Ala Leu Ser Ser Val Leu Asp 115 120 125  
 Gln Glu Glu Ala Met Glu Ile Lys Glu His Phe Pro Glu Glu Gly Ser 130 135 140  
 Ser Gly Ser Glu Val Glu Glu Ile Pro Glu Thr Pro Cys Glu Ser Gln 145 150 155 160  
 55 Gly Glu Glu Leu Lys Glu Glu Asn Met Glu Ser Val Pro Leu His Leu 165 170 175  
 Ser Leu Thr Glu Thr Gln Ser Gln Gly Leu Cys Leu Arg Arg His Pro 180 185 190  
 60

Lys Lys Lys Lys  
 195

5

(2) INFORMATION FOR SEQ ID NO: 325:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 252 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

15 Met Gly Gly Asp Leu Val Leu Gly Leu Gly Ala Leu Arg Arg Lys 1 5 10 15  
 Arg Leu Leu Glu Gln Glu Lys Ser Leu Ala Gly Trp Ala Leu Val Leu 20 25 30  
 20 Ala Xaa Xaa Gly Ile Gly Leu Met Val Leu His Ala Glu Met Leu Trp 35 40 45  
 Phe Gly Gly Cys Ser Ala Val Asn Ala Thr Gly His Leu Ser Asp Thr 50 55 60  
 25 Leu Trp Leu Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr Gly Asp Val 65 70 75 80  
 Val Pro Gly Thr Met Trp Gly Lys Ile Val Cys Leu Cys Thr Gly Val 85 90 95  
 Met Gly Val Cys Cys Thr Ala Leu Leu Val Ala Val Val Ala Arg Lys 100 105 110  
 35 Leu Glu Phe Asn Lys Ala Glu Lys His Val His Asn Phe Met Met Asp 115 120 125  
 Ile Gln Tyr Thr Lys Glu Met Lys Glu Ser Ala Ala Arg Val Leu Gln 130 135 140  
 40 Glu Ala Trp Met Phe Tyr Lys His Thr Arg Arg Lys Glu Ser His Ala 145 150 155 160  
 Ala Arg Xaa His Gln Arg Xaa Leu Leu Ala Ala Ile Asn Ala Phe Arg 165 170 175  
 45 Gln Val Arg Leu Lys His Arg Lys Leu Arg Glu Gln Val Asn Ser Met 180 185 190  
 50 Val Asp Ile Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln Asn 195 200 205  
 Leu Ser Ser Ser His Arg Ala Leu Glu Lys Gln Ile Asp Thr Leu Ala 210 215 220  
 55 Gly Lys Leu Asp Ala Leu Thr Glu Leu Ser Thr Ala Leu Gly Pro 225 230 235 240  
 Arg Gln Leu Pro Glu Pro Ser Gln Gln Ser Lys Xaa 245 250  
 60

550



(2) INFORMATION FOR SEQ ID NO: 326:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 68 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

Met Trp Arg Cys Arg Gly Lys Leu Ser Phe Pro Leu Phe Ala Val Val  
1 5 10 15

15 Ile Val Ser Cys Arg Lys Asp Gly Pro Asp Ala Ala Ala Pro Ala  
20 25 30

Val Ile Lys Asn Asn Ser His Tyr Gln Thr Ser Lys Ala Leu Gln Leu  
35 40 45

20 Gln Lys Thr Thr Gln Asn Lys Gln Ser Asn Pro Phe Ile Leu Gln Val  
50 55 60

25 Asn Lys Leu Xaa  
65

(2) INFORMATION FOR SEQ ID NO: 327:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

Met Gly Gln Gly Lys Asn Gly Phe Gly Gly Phe Val His Thr Ala Asp  
1 5 10 15

40 Ala Cys Trp Gln Gly Val His Ser Gln Pro Val Cys Arg Thr Val His  
20 25 30

Thr Val His Thr Cys His His Gln Ala Phe Leu Val Leu Ile Gly Trp  
35 40 45

45 Ser Lys Ser Gly Lys Gln Arg Lys Gln Ala Phe Leu Thr Ala Ile Ile  
50 55 60

50 Leu Asn Ser Arg Ser Ile His Ile Ser Cys Ser Trp Pro Pro Ser Pro  
65 70 75 80

Val Pro Gln Xaa

(2) INFORMATION FOR SEQ ID NO: 328:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

5 Met Leu Leu Ile Asn Leu Leu Trp Leu Val Thr Met Ile Lys Ser Val  
1 5 10 15

Ile Asn Asn Asn Ile Ile Leu Phe Leu Lys Lys Lys Ser Leu Phe Phe  
20 25 30

10 Ile Asp Ser Val  
35

15 (2) INFORMATION FOR SEQ ID NO: 329:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 63 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

25 Met Thr Phe Pro Phe Gln Lys Lys Ile Val Ala Phe Ser Ala Phe Tyr  
1 5 10 15

Leu Ile Pro Gly Gln Ser Arg Leu Ala Pro Thr Phe Asn Pro Ser Ala  
20 25 30

30 Asp Met Thr Val Ile Leu Arg Gly Arg Ala Gln His Lys Thr Ala Met  
35 40 45

Leu Gln Ser Tyr Asn Trp Lys Val Ser Cys Gln Leu Arg Gln Xaa  
50 55 60

35 (2) INFORMATION FOR SEQ ID NO: 330:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

45 Met His Ser Lys Gly Ser Ser Leu Leu Leu Phe Leu Pro Gln Leu Ile  
1 5 10 15

50 Leu Ile Leu Pro Val Cys Ala His Leu His Gln Gln Leu Asn Cys  
20 25 30

Phe His Arg  
35

(2) INFORMATION FOR SEQ ID NO: 331:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

553

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:  
5 Met Gly Ala Leu Val Leu Leu Cys Leu Leu Val Gly Val Gln Gln 15  
1  
Ser Gly Ser Val Trp Asp Ser 20

10

(2) INFORMATION FOR SEQ ID NO: 332:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 332:  
20 Met Gln Ser Ala Glu Ile Leu Ser Trp Trp Asp Val Leu Leu His Asp Phe 15  
1  
Leu Phe Ser Leu Phe Leu Trp Pro Ala Phe Glu Asp Arg Ala Leu Leu 30  
25  
Ile Phe Thr Leu Asn Gln Ile Val 40  
35

30

(2) INFORMATION FOR SEQ ID NO: 333:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 111 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 333:  
40 Met Gln Ser Leu Val Gln Trp Gly Leu Asp Ser Tyr Asp Tyr Leu Gln 15  
1  
Asn Ala Pro Pro Gly Phe Phe Pro Arg Leu Gly Val Ile Gly Phe Ala 30  
45  
Gly Leu Ile Gly Leu Leu Ala Arg Gly Ser Lys Ile Lys Lys Leu 45  
35  
Val Tyr Pro Pro Gly Phe Met Gly Leu Ala Ala Ser Leu Tyr Tyr Pro 60  
50  
Gln Gln Ala Ile Val Phe Ala Gln Val Ser Gly Glu Arg Leu Tyr Asp 80  
65  
Trp Gly Leu Arg Gly Tyr Ile Val Ile Glu Asp Leu Trp Lys Glu Asn 95  
85  
Phe Gln Lys Pro Gly Asn Val Lys Asn Ser Pro Gly Thr Lys Xaa 110  
100

60

(2) INFORMATION FOR SEQ ID NO: 334:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 106 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 334:  
10 Met Ala Pro Ser Leu Leu Leu Ala Pro Leu Cys Ser Leu Glu Ala 15  
1  
Val Leu Ser Ser Pro Leu Glu Lys Gln Cys Gln Leu Pro Gly Ile Ile Phe 30  
15  
Cys Gln Leu Gln Leu Pro Cys Pro Leu Leu Ser Ala Gln Leu Leu 45  
35  
Lys Gly Ile Val Xaa Pro Arg Cys Pro Ala Ser Leu Pro Gln Pro Pro 60  
50  
His Pro Ala Pro Ser Trp His Leu Pro Leu His Cys Thr Glu Arg Xaa 80  
65  
Pro His His Leu Pro Leu Gln Gly Gly Ser Ser Asn Met Glu Glu Xaa 95  
85  
Asn Tyr Arg Gly Tyr Xaa Asp Ala Gln Leu 105  
100

35

(2) INFORMATION FOR SEQ ID NO: 335:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 50 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 335:  
40 Met Thr Thr Cys Leu Phe Gly Leu Leu Ser Cys Glu Met Ser Ala Gln 15  
1  
Val Ser Gln Lys Ser Cys Val Tyr Asp Glu Ser Glu Cys Phe Ser Ser 30  
20  
Val Gly Gln Leu Leu Ala Leu Leu Ile Leu Val Tyr Val Leu Pro Ser 45  
35  
Ile Xaa 50  
50

55

(2) INFORMATION FOR SEQ ID NO: 336:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 48 amino acids  
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 336:

5 Met Leu Trp Lys Cys Ser Gln Asn Ile Ala Arg Cys Leu Leu Leu Leu  
1 5 10 15  
Leu Ala Leu Val Glu Ile Lys Leu Glu Asp Leu Leu Gln Ser Gln Leu His  
20 25 30  
10 Pro Thr Trp Lys Ser Ile Pro Gly Pro Ser Pro Arg Asn Gln His Arg  
35 40 45

15

(2) INFORMATION FOR SEQ ID NO: 337:

20 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 337:

25 Met Leu Ile Pro Leu Gln Cys Leu Phe Ser Ser Asp Arg Met Leu Thr  
1 5 10 15  
Phe Leu Thr Pro Trp Gln Lys Gly Glu Lys Cys Val Leu Gly Trp Val  
20 25 30  
Thr Lys Phe Leu Ser Glu Ile Ser Xaa  
35 40

35

(2) INFORMATION FOR SEQ ID NO: 338:

40 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 76 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 338:

45 Met Thr Phe Ser Ser Leu Lys Leu Phe Val Leu Thr Thr Cys Ile Ile Lys  
1 5 10 15  
Gly Leu Glu Arg Phe Ile Ile Leu Arg Glu Val Cys Asn Gln Glu Ile  
20 25 30  
50 Gln Arg Ser Leu Ser Ser Asn Leu Val His Val Leu Leu Gln Pro Ala  
35 40 45  
Thr Phe Lys Asp Val Leu Val Thr Glu Ile Ile Cys Leu Cys Met Cys  
55 60  
Leu Tyr Ser Ile Lys Tyr Met Pro Pro Gln Lys Lys  
65 70 75

60

(2) INFORMATION FOR SEQ ID NO: 339:

5 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

10 Lys Val Tyr Ile Phe Leu Ile Phe Met Val Leu Ile Leu Pro Ser Leu  
1 5 10 15  
Gly Leu Thr Arg Tyr Met Pro Pro Xaa Ser Xaa Leu Asn Ser Glu  
20 25 30

15

(2) INFORMATION FOR SEQ ID NO: 340:

20 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

25 Met Ala Lys Ile Ser Pro Phe Glu Val Val Lys Arg Thr Ser Val Pro  
1 5 10 15  
Val Leu Val Gly Leu Val Ile Val Ile Val Ala Thr Glu Met Val  
20 25 30  
Pro Gly Thr Ala Ala Val Thr Gly Lys  
35 40

35

(2) INFORMATION FOR SEQ ID NO: 341:

40 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

45 Met Arg Leu Phe Phe Ile Gly Phe Leu Leu Leu Phe Ser Phe Gly Leu  
1 5 10 15  
Leu Arg Gln Pro Ser Leu Ser Ala Glu His  
20 25

50

(2) INFORMATION FOR SEQ ID NO: 342:

55 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

60

Met Val Phe Ser Val Ser Ala Leu Ala Leu Leu Met Leu Leu  
1 5 10 15

Arg Ser Ser Asp Leu Ala Lys Lys Thr Glu  
20 25

(2) INFORMATION FOR SEQ ID NO: 343:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 157 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

Met Ser Leu Glu Phe Tyr Gln Lys Lys Lys Ser Arg Trp Pro Phe Ser  
1 5 10 15

Asp Glu Cys Ile Pro Trp Glu Val Trp Thr Val Lys Val His Val Val  
20 25 30

Ala Leu Ala Thr Glu Gln Glu Arg Gln Ile Cys Arg Glu Lys Val Gly  
35 40 45

Glu Lys Leu Cys Glu Lys Ile Ile Asn Ile Val Glu Val Met Asn Arg  
50 55 60

His Glu Tyr Leu Pro Lys Met Pro Thr Gln Ser Glu Val Asp Asn Val  
65 70 75 80

Phe Asp Thr Gly Leu Arg Asp Val Gln Pro Tyr Leu Tyr Lys Ile Ser  
85 90 95

Phe Gln Ile Thr Asp Ala Leu Gly Thr Ser Val Thr Thr Thr Met Arg  
100 105 110

Arg Leu Ile Lys Asp Thr Leu Pro Ser Glu Arg Arg Trp Ile Ser Gly  
115 120 125

Ser Ser Leu Met Ala Pro Arg Pro Trp Leu Leu Gly Ile Ala Leu Leu  
130 135 140

Gly Leu Trp Ala Leu Glu Pro Ala Leu Gly His Trp Xaa  
145 150 155

(2) INFORMATION FOR SEQ ID NO: 344:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 520 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

Met Phe Leu Leu Pro Leu Pro Ala Ala Gly Arg Val Val Val Arg Arg  
1 5 10 15

Leu Ala Val Arg Arg Phe Gly Ser Arg Ser Leu Ser Thr Ala Asp Met

Thr Lys Gly Leu Val Leu Gly Ile Tyr Ser Lys Glu Lys Glu Asp Asp  
35 40 45

Val Pro Gln Phe Thr Ser Ala Gly Glu Asn Phe Asp Lys Leu Leu Ala  
50 55 60

Gly Lys Leu Arg Glu Thr Leu Asn Ile Ser Gly Pro Pro Leu Lys Ala  
65 70 75 80

Gly Lys Thr Arg Thr Phe Tyr Gly Leu His Gln Asp Phe Pro Ser Val  
85 90 95

Val Leu Val Gly Leu Gly Lys Lys Ala Ala Gly Ile Asp Glu Gln Glu  
100 105 110

Asn Trp His Glu Gly Lys Glu Asn Ile Arg Ala Ala Val Ala Gly  
115 120 125

Cys Arg Gln Ile Gln Asp Leu Glu Ser Ser Val Glu Val Asp Pro  
130 135 140

Cys Gly Asp Ala Gln Ala Ala Glu Gly Ala Val Leu Gly Leu Tyr  
145 150 155 160

Glu Tyr Asp Asp Leu Lys Gln Lys Lys Lys Met Ala Val Ser Ala Lys  
165 170 175

Leu Tyr Gly Ser Gly Asp Gln Glu Ala Trp Gln Lys Gly Val Leu Phe  
180 185 190

Ala Ser Gly Gln Asn Leu Ala Arg Gln Leu Met Glu Thr Pro Ala Asn  
195 200 205

Glu Met Thr Pro Thr Arg Phe Ala Glu Ile Ile Glu Lys Asn Leu Lys  
210 215 220

Ser Ala Ser Ser Lys Thr Glu Val His Ile Arg Pro Lys Ser Trp Ile  
225 230 235 240

Glu Glu Gln Ala Met Gly Ser Phe Leu Ser Val Ala Lys Gly Ser Asp  
245 250 255

Glu Pro Pro Val Phe Leu Glu Ile His Tyr Lys Gly Ser Pro Asn Ala  
260 265 270

Asn Glu Pro Pro Leu Val Phe Val Gly Lys Gly Ile Thr Phe Asp Ser  
275 280 285

Gly Gly Ile Ser Ile Lys Ala Ser Ala Asn Met Asp Leu Met Arg Ala  
290 295 300

Asp Met Gly Gly Ala Ala Thr Ile Cys Ser Ala Ile Val Ser Ala Ala  
305 310 315 320

Lys Leu Asn Leu Pro Ile Asn Ile Ile Gly Leu Ala Pro Leu Cys Glu  
325 330 335

Asn Met Pro Ser Gly Lys Ala Asn Lys Pro Gly Asp Val Val Arg Ala

340 345 350  
Lys Asn Gly Lys Thr Ile Gln Val Asp Asn Thr Asp Ala Glu Gly Arg  
355 360 365  
5 Leu Ile Leu Ala Asp Ala Leu Cys Tyr Ala His Thr Phe Asn Pro Lys  
370 375 380  
10 Xaa Ile Leu Asn Ala Ala Thr Leu Thr Gly Ala Met Asp Val Ala Leu  
385 390 395 400  
Gly Ser Gly Ala Thr Gly Val Phe Thr Asn Ser Ser Trp Leu Trp Asn  
405 410 415  
15 Lys Leu Phe Glu Ala Ser Ile Glu Thr Gly Asp Arg Val Trp Arg Met  
420 425 430  
Pro Leu Phe Glu His Tyr Thr Arg Gln Val Val Asp Cys Gln Leu Ala  
435 440 445  
20 Asp Val Asn Asn Ile Gly Lys Tyr Arg Ser Ala Gly Ala Cys Thr Ala  
450 455 460  
Ala Ala Phe Leu Lys Glu Phe Val Thr His Pro Lys Trp Ala His Leu  
465 470 475 480  
25 Asp Ile Ala Gly Val Met Thr Asn Lys Asp Glu Val Pro Tyr Leu Arg  
485 490 495  
30 Lys Gly Met Thr Gly Arg Pro Thr Arg Thr Leu Ile Glu Phe Leu Leu  
500 505 510  
Arg Phe Ser Gln Asp Asn Ala Xaa  
515 520  
35 (2) INFORMATION FOR SEQ ID NO: 345:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 345:  
40 Thr Ile Leu Phe Leu Phe Leu Gln Leu Ser Ala Leu Arg Leu Ile Val  
1 5 10 15  
45 Gly Lys Asp Ser Ile Asp Ile Asp Ile Ser Ser Arg Arg Arg Glu Asp  
20 25 30  
50 Gln Ser Leu Arg Leu Asn Ala  
35  
55 (2) INFORMATION FOR SEQ ID NO: 346:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 234-amino acids  
60

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 346:  
5 Met Thr Ser Glu Leu Asp Ile Phe Val Gly Asn Thr Thr Leu Ile Asp  
1 5 10 15  
Glu Asp Val Tyr Arg Leu Trp Leu Asp Gly Tyr Ser Val Thr Asp Ala  
20 25 30  
10 Val Ala Leu Arg Val Arg Ser Gly Ile Leu Glu Gln Thr Gly Ala Thr  
35 40 45  
15 Ala Ala Val Leu Gln Ser Asp Thr Met Asp His Tyr Arg Thr Phe His  
50 55 60  
Met Leu Glu Arg Leu Leu His Ala Pro Pro Lys Leu Leu His Gln Leu  
65 70 75 80  
20 Ile Phe Gln Ile Pro Pro Ser Arg Gln Ala Leu Leu Ile Glu Arg Tyr  
85 90 95  
Tyr Ala Phe Asp Glu Ala Phe Val Arg Glu Val Leu Gly Lys Lys Leu  
100 105 110  
25 Ser Lys Gly Thr Lys Lys Asp Leu Asp Asp Ile Ser Thr Lys Thr Gly  
115 120 125  
Ile Thr Leu Lys Ser Cys Arg Arg Gln Phe Asp Asn Phe Lys Arg Val  
130 135 140  
30 Phe Lys Val Val Glu Glu Met Arg Gly Ser Leu Val Asp Asn Ile Gln  
145 150 155 160  
35 Gln His Phe Leu Leu Ser Asp Arg Leu Ala Arg Asp Tyr Ala Ala Ile  
165 170 175  
Val Phe Phe Ala Asn Asn Arg Phe Glu Thr Gly Lys Lys Lys Leu Gln  
180 185 190  
40 Tyr Leu Ser Phe Gly Asp Phe Ala Phe Cys Ala Glu Leu Met Ile Gln  
195 200 205  
Asn Trp Thr Leu Gly Pro Val Asp Ser Gln Met Asp Asp Met Asp Met  
210 215 220  
45 Asp Leu Asp Arg Asn Phe Ser Arg Thr Xaa  
225 230  
50 (2) INFORMATION FOR SEQ ID NO: 347:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 169 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 347:  
55 Met Ala Ala Ala Val Ala Gly Met Leu Arg Gly Gly Leu Leu Pro Gln  
60

1 5 10 15  
 Ala Gly Arg Leu Pro Thr Leu Gln Thr Val Arg Tyr Gly Ser Lys Ala 30  
 20 25 30  
 5 Val Thr Arg His Arg Arg Val Met His Phe Gln Arg Gln Lys Leu Met 45  
 35 40 45  
 10 Ala Val Thr Glu Tyr Ile Pro Pro Lys Pro Ala Ile His Pro Ser Cys 60  
 50 55 60  
 Leu Pro Ser Pro Pro Ser Pro Gln Glu Glu Ile Gly Leu Ile Arg 75  
 65 70 80  
 15 Leu Leu Arg Arg Glu Ile Ala Val Phe Gln Asp Asn Arg Met Ile 95  
 85 90 95  
 Ala Val Cys Gln Asn Val Ala Leu Ser Ala Glu Asp Lys Leu Leu Ile 110  
 100 105 110  
 20 Ala Thr Pro Ala Ala Glu Thr Gln Asp Pro Asp Glu Gly Leu Pro Gln 125  
 115 120 125  
 25 Pro Gly Pro Glu Ser Pro Ser Trp Arg Ile Pro Ser Thr Lys Ile Cys 140  
 130 135 140  
 Cys Pro Phe Leu Trp Gly Thr Thr Cys Cys Trp Ser Val Lys Ser Pro 155  
 145 150 160  
 30 Arg Ser Arg Arg Trp Tyr Gly Ser Xaa 165  
 160 165

35 (2) INFORMATION FOR SEQ ID NO: 348:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

Met Lys Arg Ser Phe Leu Leu Pro Leu Leu Leu Val Gly Phe Leu Asp 15  
 1 5 10 15  
 45 Thr Ala His Leu Ile Leu Leu Glu Thr Leu Ser Val Cys Leu Trp Leu 30  
 20 25 30  
 Pro Ser Leu Ile Asp Ser Arg Cys Val Met Ser 40  
 35 40 45

(2) INFORMATION FOR SEQ ID NO: 349:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 78 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

Met Lys Glu Gly Pro Cys Lys Arg His His Tyr Tyr Gln Asn Cys 15  
 1 5 10 15  
 5 Gly Ala Lys Leu Leu Val Ser Leu Phe Gly Glu Thr Asn Gln Ile His 30  
 20 25 30  
 Leu Leu Glu Thr Gln Val Thr Glu Lys Gly Gly Glu Arg Ile Trp 45  
 35 40 45  
 10 Glu Glu Lys Trp Arg Ile Ser Ser Thr Val Leu Phe Ile Ser Val Asn 60  
 50 55 60  
 Ser Tyr Val Glu Gly Ser Val Leu Glu Ile Lys Leu Phe Tyr 75  
 65 70 75  
 20 (2) INFORMATION FOR SEQ ID NO: 350:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

Met Ser Glu Ile Leu Ser Leu Leu Phe Cys Leu Leu Gly Pro Ala Leu 15  
 1 5 10 15  
 30 Asp Glu Arg Arg Glu Glu Lys Asp 20  
 20 25

35 (2) INFORMATION FOR SEQ ID NO: 351:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 274 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

Met Ser Ser Ala Gly Thr Ala Thr Pro Leu Glu Met Asp His Lys Leu 15  
 1 5 10 15  
 45 Thr Ser Gln Pro Gly Arg Pro Ser Phe Tyr Cys Asn Ser Arg His Ser 30  
 20 25 30  
 Ile Val Gly Ser Ser His Gln Leu Gly Phe Trp Phe Ser His Leu Glu 45  
 35 40 45  
 Ser Ser Gly Leu Lys Val Phe Gln Val Ser Leu Pro Cys Glu Cys Val 60  
 50 55 60  
 Asn Leu Pro Thr Arg Ile Ala Ser Val Val Leu Ser Leu Met Ser Leu 80  
 65 70 75  
 Leu Val Val Gly Gln Ala Pro Ala Trp Glu Gly Ser Leu Leu Arg Gly 95  
 85 90 95  
 60

Arg Pro Ala Gly Ala His Leu Cys Ala Met Xaa Val Ile Glu Gly  
100 105 110  
Leu Val Val Asp Val Gly Glu Arg Ile Leu His Gly Gln Arg Glu Val  
115 120 125  
Gly Gln Val Ser Gln Val Leu Pro Ala Leu Ser Leu Gly Leu Val Phe  
130 135 140  
Leu Cys Gln Gly Thr Val Glu Lys Val Ser Gly Ala Ala His Cys Ser  
145 150 155 160  
Ser Leu Leu Cys Cys Leu Pro Trp Gln Cys Ser Gly Gly Gly Phe Pro  
165 170 175  
Thr Xaa Arg Cys Ser Arg Pro Tyr Phe Ser Ser His Lys Gly Val Ala  
180 185 190  
Ala Thr Leu Ala Leu Thr Cys His Cys Asp Lys Val His Val Ala Gly  
195 200 205  
Leu Gly Lys Asp Trp Ala Ile Glu Gln Arg Arg Thr Cys Glu Ser  
210 215 220  
Asp Xaa Glu Xaa Xaa Pro Phe Thr Leu Ala Gly Leu Val Leu Val Leu  
225 230 235 240  
Arg Phe Cys Gln Val Val Leu Val Trp Ile Pro Gln Leu Gly Asp Lys  
245 250 255  
His Trp Arg Gly Met Thr Arg Leu Gly Arg Val Ser Leu Thr Ser Ser  
260 265 270  
Ile Xaa  
35  
(2) INFORMATION FOR SEQ ID NO: 352:  
40 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
45 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 352:  
Met Ile Phe Thr Ser Val Thr Lys Gly Ile Leu Leu Ile Ala Leu Trp  
1 5 10 15  
Val Pro Leu Phe His Phe Met Leu Ile Asp Ser Ile Leu Gly Pro Ser  
20 25 30  
Arg Leu Leu Thr Asp Gly Val Pro Phe Asn Pro Trp His Val Xaa  
35 40 45  
(2) INFORMATION FOR SEQ ID NO: 353:  
60 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 353:  
Met Lys Thr  
1  
10 (2) INFORMATION FOR SEQ ID NO: 354:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 52 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 354:  
Met Ser Ile Ser Gly Thr Asp Gly Leu Ile Leu Leu Leu Val Gly Leu  
20 1 5 10 15  
Glu Ala Xaa Val Arg Ser Ser Lys Lys Trp Ile Pro Lys Ala Leu Xaa  
20 25 30  
Val Thr Gln Ala Lys Trp Asn Ser Trp Pro Ser Arg Arg Asn Ala Gly  
35 40 45  
Phe Ala Leu His  
50  
30 (2) INFORMATION FOR SEQ ID NO: 355:  
35 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 132 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
40 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 355:  
Met Glu His Cys Leu Tyr His Ser Val His Gly Ile Asn Pro Tyr Ile  
1 5 10 15  
His Lys Asn Thr His Pro Ser Ile Asn Ile Tyr Met Val Trp Asp Glu  
20 25 30  
Gln Val Asn Ser Phe Glu Arg Glu Phe Val Pro Phe Phe Leu Ile  
35 40 45  
Ile Leu Leu Asn Cys Cys Gln Leu Ser Asn Lys Gln Thr Glu Lys Leu  
50 55 60  
Phe Gly Lys Thr Leu His Thr Pro Phe Leu Ser Ser Ala Leu Lys Tyr  
65 70 75 80  
Arg Leu Asn Thr His Ile Leu Pro Val Phe Ser Tyr Ser Asp Ser Ile  
85 90 95  
Leu Thr Cys His Leu Ile Leu Ala Ser Tyr Phe Ser His Val Tyr Leu  
100 105 110

565

Pro Val Thr Cys Ile Cys Tyr Leu Asn Arg Lys Lys Asn Ile Gln Lys  
115 120

5 Lys Lys Asn Xaa  
130

10 (2) INFORMATION FOR SEQ ID NO: 356:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 204 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

Met Gly Ser Arg Asp His Leu Phe Lys Val Leu Val Val Gly Asp Ala  
1 5 10 15

20 Ala Val Gly Lys Thr Ser Leu Val Gln Asp Tyr Ser Gln Asp Ser Phe  
20 25 30

25 Ser Lys His Tyr Lys Ser Thr Val Gly Val Asp Phe Ala Leu Lys Val  
35 40 45

Leu Gln Trp Ser Asp Tyr Glu Ile Val Arg Leu Gln Leu Trp Asp Ile  
50 55 60

30 Ala Gly Gln Glu Arg Phe Thr Ser Met Thr Arg Leu Tyr Tyr Arg Asp  
65 70 75 80

Ala Ser Ala Cys Val Ile Met Phe Asp Val Thr Asn Ala Thr Thr Phe  
85 90 95

35 Ser Asn Ser Gln Arg Trp Lys Gln Asp Leu Asp Ser Lys Leu Thr Leu  
100 105 110

40 Pro Asn Gly Glu Pro Val Pro Cys Leu Leu Ala Asn Lys Cys Asp  
115 120 125

Leu Ser Pro Trp Ala Val Ser Arg Asp Gln Ile Asp Arg Phe Ser Lys  
130 135 140

45 Glu Asn Gly Phe Thr Gly Trp Thr Glu Thr Ser Val Lys Glu Asn Lys  
145 150 155 160

Asn Ile Asn Glu Ala Met Arg Val Leu Ile Glu Lys Met Met Arg Asn  
165 170 175

50 Ser Thr Glu Asp Ile Met Ser Leu Ser Thr Gln Gly Asp Tyr Ile Asn  
180 185 190

55 Leu Gln Thr Lys Ser Ser Trp Ser Cys Xaa  
195 200

(2) INFORMATION FOR SEQ ID NO: 357:

60

566

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

5

Met Ile Ser Leu Ile Phe Gln Leu Glu Glu Lys Leu Val Glu Lys  
1 5 10 15

10 Phe Phe Phe Leu Phe Phe Leu Lys Lys Gly Ser Gln Gly Ser  
20 25 30

15 Asn Leu Lys Ile Val Pro Arg His Met Arg Val Val Leu Arg Gly  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 358:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 73 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

25

Met Thr Tyr Val Thr Cys Leu His Val Cys Leu Leu Val Glu Phe Leu  
1 5 10 15

30 Asn Ser Gln Leu Thr Asn His Arg Lys Tyr Tyr Phe Leu Ser Tyr Gly  
20 25 30

Phe Trp Phe Thr Gly Leu Arg Gly Phe Ser Glu Tyr Leu Trp Pro Gln  
35 40 45

35 Gln His Thr Ser Phe His Pro Asn Arg Asn Glu Ile Asn Phe Val Ser  
50 55 60

Thr Asp Asn Arg Ile Trp Val Thr Xaa  
65 70

40

(2) INFORMATION FOR SEQ ID NO: 359:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 102 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

45

Met Ser Asp Gln Glu Ala Lys Pro Ser Thr Glu Asp Leu Gly Asp Lys  
1 5 10 15

55 Lys Glu Gly Glu Tyr Ile Lys Leu Lys Val Ile Gly Gln Asp Ser Ser  
20 25 30

Glu Ile His Phe Lys Val Lys Met Thr Thr His Leu Lys Lys Leu Lys  
35 40 45

60 Glu Ser Tyr Cys Gln Arg Gln Gly Val Pro Met Asn Ser Leu Arg Phe  
40 45



50 55 60  
Leu Phe Glu Gly Gln Arg Ile Ala Asp Asn His Thr Pro Lys Glu Leu  
65 70 75 80  
5 Gly Met Glu Glu Glu Asp Val Ile Glu Val Tyr Gln Glu Gln Thr Gly  
85 90 95  
10 Gly His Ser Thr Val Xaa  
100  
15 (2) INFORMATION FOR SEQ ID NO: 360:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 48 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 360:  
Met Gly Phe Pro Gln Trp His Leu Gly Asn His Ala Val Glu Pro Val  
1 5 10 15  
25 Thr Ser Ile Leu Leu Leu Phe Leu Leu Met Met Leu Gly Val Arg Gly  
20 25 30  
30 Leu Leu Leu Val Gly Leu Val Tyr Leu Val Ser His Leu Ser Gln Arg  
35 40 45  
35 (2) INFORMATION FOR SEQ ID NO: 361:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 179 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
40 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 361:  
Met Ser Ala Glu Val Lys Val Thr Gly Gln Asn Gln Glu Gln Phe Leu  
1 5 10 15  
45 Leu Leu Ala Lys Ser Ala Lys Gly Ala Ala Leu Ala Thr Leu Ile His  
20 25 30  
50 Gln Val Leu Glu Ala Pro Gly Val Tyr Val Phe Gly Glu Leu Leu Asp  
35 40 45  
55 Met Pro Asn Val Arg Glu Leu Ala Glu Ser Asp Phe Ala Ser Thr Phe  
50 55 60  
60 Arg Leu Leu Thr Val Phe Ala Tyr Gly Thr Tyr Ala Asp Tyr Leu Ala  
65 70 75 80  
Glu Ala Arg Asn Leu Pro Leu Thr Glu Ala Gln Lys Asn Lys Leu  
85 90 95

Arg His Leu Ser Val Val Thr Leu Ala Ala Lys Val Lys Cys Ile Pro  
100 105 110  
5 Tyr Ala Val Leu Leu Glu Ala Leu Ala Leu Arg Asn Val Arg Gln Leu  
115 120 125  
10 Glu Asp Leu Val Ile Glu Ala Val Tyr Ala Asp Val Leu Arg Gly Ser  
130 135 140  
145 Leu Asp Gln Arg Asn Gln Arg Leu Glu Val Asp Tyr Ser Ile Gly Arg  
150 155 160  
15 Asp Ile Gln Arg Gln Asp Leu Ser Ala Ile Ala Arg Thr Leu Xaa Lys  
165 170 175  
Asn His Xaa  
20 (2) INFORMATION FOR SEQ ID NO: 362:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 362:  
Met Lys Ser Ser Leu Phe Phe Phe Leu Ala His Phe Ile His  
1 5 10 15  
30 Ser His Asp Leu Pro Gly Leu Cys Arg  
20 25  
35 (2) INFORMATION FOR SEQ ID NO: 363:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 224 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
40 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 363:  
Met Lys Phe Ala Ala Ser Gly Xaa Phe Leu His His Met Ala Gly Leu  
1 5 10 15  
45 Ser Ser Lys Leu Ser Met Ser Lys Ala Leu Pro Leu Thr Lys Val  
20 25 30  
50 Val Gln Asn Asp Ala Tyr Thr Ala Pro Ala Leu Pro Ser Ser Ile Arg  
35 40 45  
55 Thr Lys Ala Leu Thr Asn Met Ser Arg Thr Leu Val Asn Lys Glu Glu  
50 55 60  
60 Pro Pro Lys Glu Leu Pro Ala Ala Glu Pro Val Leu Ser Pro Leu Glu  
65 70 75 80

Gly Thr Lys Met Thr Val Asn Asn Leu His Pro Arg Val Thr Glu Glu 95  
85

5 Asp Ile Val Glu Leu Phe Cys Val Cys Gly Ala Leu Lys Arg Ala Arg 110  
100

Leu Val His Pro Gly Val Ala Glu Val Val Phe Val Lys Lys Asp Asp 125  
115

10 Ala Ile Thr Ala Tyr Lys Lys Tyr Asn Asn Arg Cys Leu Asp Gly Gln 140  
130

Pro Met Lys Cys Asn Leu His Met Asn Gly Asn Val Ile Thr Ser Asp 160  
145

15 Gln Pro Ile Leu Leu Arg Leu Ser Asp Ser Pro Ser Met Lys Lys Glu 175  
165

Ser Glu Leu Pro Arg Arg Val Asn Ser Ala Ser Ser Ser Asn Pro Pro 190  
180

Ala Glu Val Asp Pro Asp Thr Ile Leu Lys Ala Leu Phe Lys Ser Ser 205  
195

25 Gly Ala Ser Xaa Thr Thr Gln Pro Thr Glu Phe Lys Ile Lys Leu Xaa 220  
210

30

(2) INFORMATION FOR SEQ ID NO: 364:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 349 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 364:

40 Met Ser Lys Asn Cys Ile Lys Leu Leu Cys Glu Asp Pro Val Phe Ala 15  
1

Glu Tyr Ile Lys Cys Ile Leu Met Asp Glu Arg Thr Phe Leu Asn Asn 30  
20

Asn Ile Val Tyr Thr Phe Met Thr His Phe Leu Leu Lys Val Gln Ser 45  
35

50 Gln Val Phe Ser Glu Ala Asn Cys Ala Asn Leu Ile Ser Thr Leu Ile 60  
50

Thr Asn Leu Ile Ser Gln Tyr Gln Asn Leu Gln Ser Asp Phe Ser Asn 80  
65

55 Arg Val Glu Ile Ser Lys Ala Ser Ala Ser Leu Asn Gly Asp Leu Arg 95  
85

Ala Leu Ala Leu Leu Ser Val His Thr Pro Lys Gln Leu Asn Pro 110  
100

Ala Leu Ile Pro Thr Leu Gln Glu Leu Ser Lys Cys Arg Thr Cys 125  
115

5 Leu Gln Gln Arg Asn Ser Leu Gln Glu Gln Ala Lys Glu Arg Lys 140  
130

Thr Lys Asp Asp Glu Gly Ala Thr Pro Ile Lys Arg Arg Val Ser 155  
145

10 Ser Asp Glu Glu His Thr Val Asp Ser Cys Ile Ser Asp Met Lys Thr 175  
165

Glu Thr Arg Glu Val Leu Thr Pro Thr Ser Thr Ser Asp Asn Glu Thr 190  
180

15 Arg Asp Ser Ser Ile Ile Asp Pro Gly Thr Glu Gln Asp Leu Pro Ser 205  
195

20 Pro Glu Asn Ser Ser Val Lys Glu Tyr Arg Met Glu Val Pro Ser Ser 220  
210

Phe Ser Glu Asp Met Ser Asn Ile Arg Ser Gln His Ala Glu Glu Gln 240  
225

25 Ser Asn Asn Gly Arg Tyr Asp Cys Lys Glu Phe Lys Asp Leu His 255  
245

Cys Ser Lys Asp Ser Thr Leu Ala Glu Glu Ser Glu Phe Pro Ser 270  
260

30 Thr Ser Ile Ser Ala Val Leu Ser Asp Leu Ala Asp Leu Arg Ser Cys 285  
275

35 Asp Gly Gln Ala Leu Pro Ser Gln Asp Pro Glu Val Ala Leu Ser Leu 300  
290

Ser Cys Gly His Ser Arg Gly Leu Phe Ser His Met Gln Gln His Asp 315  
305

40 Ile Leu Asp Thr Leu Cys Arg Thr Ile Glu Ser Thr Ile His Val Val 335  
325

Thr Arg Ile Ser Gly Lys Gly Asn Gln Ala Ala Ser Xaa 345  
340

(2) INFORMATION FOR SEQ ID NO: 365:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 467 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 365:

50 Met Leu His Gln Asp His Ile Thr Phe Ala Met Leu Leu Ala Arg Ile 15  
1

55 Lys Leu Lys Gly Thr Val Gly Glu Pro Thr Tyr Asp Ala Glu Phe Gln 10

20 25 30  
 His Phe Leu Arg Gly Asn Glu Ile Val Leu Ser Ala Gly Ser Thr Pro  
 35 40 45  
 5 Arg Ile Glu Gly Leu Thr Val Glu Glu Ala Glu Ala Val Val Arg Leu  
 50 55 60  
 10 Ser Cys Leu Pro Ala Phe Lys Asp Leu Ile Ala Lys Val Glu Ala Asp  
 65 70 75 80  
 Glu Glu Phe Gly Ile Trp Leu Asp Ser Ser Pro Glu Glu Thr Val  
 85 90 95  
 15 Pro Tyr Leu Trp Ser Glu Glu Thr Pro Ala Thr Pro Ile Gly Glu Ala  
 100 105 110  
 Ile His Arg Leu Leu Leu Ile Glu Ala Phe Arg Pro Asp Arg Leu Leu  
 115 120 125  
 20 Ala Met Ala His Met Phe Val Ser Thr Asn Leu Leu Gly Glu Ser Phe Met  
 130 135 140  
 Ser Ile Met Glu Glu Pro Leu Asp Leu Thr His Ile Val Xaa Thr Glu  
 145 150 155 160  
 Val Lys Pro Asn Thr Pro Val Leu Met Cys Ser Val Pro Gly Tyr Asp  
 165 170 175  
 30 Ala Ser Gly His Val Glu Asp Leu Ala Ala Glu Glu Asn Thr Glu Ile  
 180 185 190  
 Thr Ser Ile Ala Ile Gly Ser Ala Glu Gly Phe Asn Glu Ala Asp Lys  
 195 200 205  
 35 Ala Ile Asn Thr Ala Val Lys Ser Gly Arg Trp Val Met Leu Lys Asn  
 210 215 220  
 Val His Leu Ala Pro Gly Trp Leu Met Glu Leu Glu Lys Lys Leu His  
 225 230 235 240  
 Ser Leu Glu Pro His Ala Cys Phe Arg Leu Phe Leu Thr Met Glu Ile  
 245 250 255  
 45 Asn Pro Lys Val Pro Val Asn Leu Leu Arg Ala Gly Arg Ile Phe Val  
 260 265 270  
 Phe Glu Pro Pro Gly Xaa Lys Ala Asn Met Leu Arg Thr Phe Ser  
 275 280 285  
 50 Ser Ile Pro Val Ser Arg Ile Cys Lys Ser Pro Asn Glu Arg Ala Arg  
 290 295 300  
 Leu Tyr Phe Leu Leu Ala Trp Phe His Ala Ile Ile Glu Glu Arg Leu  
 305 310 315 320  
 Arg Tyr Ala Pro Leu Gly Trp Ser Lys Tyr Tyr Glu Phe Gly Glu Ser  
 325 330 335  
 60 Asp Leu Arg Ser Xaa Cys Asp Thr Val Asp Thr Trp Leu Asp Asp Thr

340 345 350  
 Ala Lys Gly Arg Glu Asn Ile Ser Pro Asp Lys Ile Pro Trp Ser Ala  
 355 360 365  
 5 Leu Lys Thr Leu Met Ala Glu Ser Ile Tyr Gly Gly Arg Val Asp Asn  
 370 375 380  
 Glu Phe Asp Glu Arg Leu Leu Asn Thr Phe Leu Glu Arg Leu Phe Thr  
 385 390 395 400  
 Thr Arg Ser Phe Asp Ser Glu Phe Lys Leu Ala Cys Lys Val Asp Gly  
 405 410 415  
 15 His Lys Asp Ile Glu Met Pro Asp Gly Met Glu Ala Arg Gly Val Cys  
 420 425 430  
 Ala Val Gly Gly Val Ala Pro Arg His Pro Asp Ala Leu Leu Ala Gly  
 435 440 445  
 20 Pro Ala Glu Glu Arg Arg Glu Ser Pro Pro Tyr His Thr Gly Cys Gly  
 450 455 460  
 His Asp Glu  
 465  
 25  
 30 (2) INFORMATION FOR SEQ ID NO: 366:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 152 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 366:  
 Met Ala Asp Glu Ala Thr Arg Arg Val Val Ser Glu Ile Pro Val Leu  
 1 5 10 15  
 Lys Thr Asn Ala Gly Pro Arg Asp Arg Glu Leu Trp Val Glu Arg Leu  
 20 25 30  
 Lys Glu Glu Tyr Glu Ser Leu Ile Arg Tyr Val Glu Asn Asn Lys Asn  
 35 40 45  
 Ala Asp Asn Asp Trp Phe Arg Leu Glu Ser Asn Lys Glu Gly Thr Arg  
 50 55 60  
 Trp Phe Gly Lys Cys Trp Tyr Ile His Asp Leu Leu Lys Tyr Glu Phe  
 65 70 75 80  
 Asp Ile Glu Phe Asp Ile Pro Ile Thr Tyr Pro Thr Thr Ala Pro Glu  
 85 90 95  
 55 Ile Ala Val Pro Glu Leu Asp Gly Lys Thr Ala Lys Met Tyr Arg Gly  
 100 105 110  
 Gly Lys Ile Cys Leu Thr Asp His Phe Lys Pro Leu Trp Gly Glu Glu  
 115 120 125  
 60

Cys Ala Gln Ile Trp Thr Ser Ser Ser His Gly Ser Gly Ala Gly Ser  
130 135 140

Met Xaa Gly Ser Gly Asn Pro Xaa  
145 150

10 (2) INFORMATION FOR SEQ ID NO: 367:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 373 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 367:

Met Tyr Asp Gly Thr Lys Glu Val Pro Met Asn Pro Val Lys Ile Tyr  
1 5 10 15

20 Gln Val Cys Asp Ile Pro Gln Pro Gln Gly Ser Ile Ile Asn Pro Gly  
20 25 30

25 Ser Thr Gly Ser Ala Pro Trp Asp Glu Lys Asp Asn Asp Val Asp Glu  
35 40 45

Glu Asp Glu Glu Asp Glu Leu Asp Gln Ser Gln His His Val Pro Ile  
50 55 60

30 Gln Asp Thr Phe Pro Phe Leu Asn Ile Asn Gly Ser Pro Met Ala Pro  
65 70 75 80

Ala Ser Val Gly Asn Cys Ser Val Gly Asn Cys Ser Pro Glu Ala Val  
85 90 95

35 Trp Pro Lys Thr Glu Pro Leu Glu Met Glu Val Pro Gln Ala Pro Ile  
100 105 110

Gln Pro Phe Tyr Ser Ser Pro Glu Leu Trp Ile Ser Ser Leu Pro Met  
115 120 125

40 Thr Asp Leu Asp Ile Lys Phe Gln Tyr Arg Gly Lys Glu Tyr Gly Gln  
130 135 140

45 Thr Met Thr Val Ser Asn Pro Gln Gly Cys Arg Leu Phe Tyr Gly Asp  
145 150 155 160

Leu Gly Pro Met Pro Asp Gln Glu Leu Phe Gly Pro Val Xaa Leu  
165 170 175

50 Glu Gln Val Lys Phe Pro Gly Pro Glu His Ile Thr Asn Glu Lys Gln  
180 185 190

Lys Leu Phe Thr Ser Lys Leu Leu Asp Val Met Asp Arg Gly Leu Ile  
195 200 205

55 Leu Glu Val Ser Gly His Ala Ile Tyr Ala Ile Arg Leu Cys Gln Cys  
210 215 220

60 Lys Val Tyr Trp Ser Gly Pro Cys Ala Pro Ser Leu Val Ala Pro Asn  
225 230 235 240

Leu Ile Glu Arg Gln Lys Lys Val Lys Leu Phe Cys Leu Glu Thr Phe  
245 250 255

5 Leu Ser Asp Leu Ile Ala His Gln Lys Gly Gln Ile Glu Lys Gln Pro  
260 265 270

10 Pro Phe Glu Ile Tyr Leu Cys Phe Gly Glu Glu Trp Pro Asp Gly Lys  
275 280 285

Pro Leu Glu Arg Lys Leu Ile Leu Val Gln Val Ile Pro Val Val Ala  
290 295 300

15 Arg Met Ile Tyr Glu Met Phe Ser Gly Asp Phe Thr Arg Ser Phe Asp  
305 310 315 320

Ser Gly Ser Val Arg Leu Gln Ile Ser Thr Pro Asp Ile Lys Asp Asn  
325 330 335

20 Ile Val Ala Gln Leu Lys Gln Leu Tyr Arg Ile Leu Gln Thr Gln Glu  
340 345 350

25 Ser Trp Gln Pro Met Gln Pro Thr Pro Ser Met Gln Leu Pro Pro Ala  
355 360 365

Leu Pro Pro Gln Xaa  
370

(2) INFORMATION FOR SEQ ID NO: 368:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 83 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 368:

Met Gly Ser Ser Val Leu Pro Phe Cys Val Cys Val Thr Ser Pro Ser  
1 5 10 15

Leu Gly Gly Arg Cys Ile Gln Gly Arg Phe Ala Ser His Ser Lys Phe  
20 25 30

45 Trp Gly Phe Gly Arg Lys Thr Ala Ser Phe Gly Ala Val Gly Glu Thr  
35 40 45

Pro Pro Asp Gln Glu Pro Gln Lys Glu Thr Glu Pro Ala Thr Ser Ser  
50 55 60

His Ala Arg Pro Trp Ala Arg Val Ile Gly Leu Arg Ile Trp Pro Gln  
65 70 75 80

Pro Asn Xaa

(2) INFORMATION FOR SEQ ID NO: 369:

60

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 369:

Met Leu Leu Ser Val Ala Ile Phe Ile Leu Leu Thr Leu Val Tyr Ala  
1 5 10 15

10 Tyr Trp Thr Met Xaa  
20

15 (2) INFORMATION FOR SEQ ID NO: 370:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 227 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 370:

Met Gly Ala Ser Ala Arg Leu Leu Arg Ala Val Ile Met Gly Ala Pro  
1 5 10 15

25 Gly Ser Gly Lys Gly Thr Val Ser Ser Arg Ile Thr Thr His Phe Glu  
20 25 30

30 Leu Lys His Leu Ser Ser Gly Asp Leu Leu Arg Asp Asn Met Leu Arg  
35 40 45

Gly Thr Glu Ile Gly Val Leu Ala Lys Ala Phe Ile Asp Glu Gly Lys  
50 55 60

35 Leu Ile Pro Asp Asp Val Met Thr Arg Leu Leu His Glu Leu Lys  
65 70 75 80

Asn Leu Thr Glu Tyr Ser Trp Leu Leu Asp Gly Phe Pro Arg Thr Leu  
85 90 95

40 Pro Glu Ala Glu Ala Leu Asp Arg Ala Tyr Glu Ile Asp Thr Val Ile  
100 105 110

45 Asn Leu Asn Val Pro Phe Glu Val Ile Lys Glu Arg Leu Thr Ala Arg  
115 120 125

Trp Ile His Pro Ala Ser Gly Arg Val Tyr Asn Ile Glu Phe Asn Pro  
130 135 140

50 Pro Lys Thr Val Gly Ile Asp Asp Leu Thr Gly Glu Pro Leu Ile Glu  
145 150 155 160

Arg Glu Asp Asp Lys Pro Glu Thr Val Ile Lys Arg Leu Lys Ala Tyr  
165 170 175

55 Glu Asp Glu Thr Lys Pro Val Leu Glu Tyr Tyr Glu Lys Lys Val  
180 185 190

60 Leu Glu Thr Phe Ser Gly Thr Glu Thr Asn Lys Ile Trp Pro Tyr Val  
195 200 205

Tyr Ala Phe Leu Glu Thr Lys Val Pro Glu Arg Ser Glu Lys Ala Ser  
210 215 220

5 Val Thr Pro  
225

10 (2) INFORMATION FOR SEQ ID NO: 371:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 371:

Met Phe Leu Asn Cys Glu Ile Leu Glu Tyr Cys Tyr Tyr Leu Thr Glu  
1 5 10 15

20 Leu Lys Ile Ser Met Gly Lys Tyr Leu Ser Ile Pro Thr Val Leu Leu  
20 25 30

25 Lys Ile Ile Arg Cys Ser Ile Thr Ala Val Ser Asp Ser Ser Thr Ser  
35 40 45

Trp Ala Ile Lys Ala Glu Leu Lys Ile Glu Asn Lys Asp Leu Asp Asn  
50 55 60

30 Lys Thr Ala Lys Gly Gly Glu Glu Ala Leu Thr Cys Thr Xaa  
65 70 75

35 (2) INFORMATION FOR SEQ ID NO: 372:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 372:

Met Arg Ala Val Phe Pro Cys Cys Pro Phe Leu Thr Leu Met Leu Pro  
1 5 10 15

45 Leu Leu Glu Cys Leu Val Gly Met Ile Met Cys Tyr Leu Gly Ile Ser  
20 25 30

50 Phe Thr Asp Thr Arg Lys Thr Ala Gly Leu Lys Lys Lys Lys  
35 40 45

Lys Xaa Xaa  
50

55 (2) INFORMATION FOR SEQ ID NO: 373:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 61 amino acids

577

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 373:

5 Met Phe Leu Met Arg Met His Leu Cys Phe Cys Lys Tyr Cys Cys Ser 15  
1  
Phe Ile Val Thr Pro Thr Ser Thr Ser Asn Thr Ala Ser Tyr Leu Trp 30  
10 Pro Trp Ile Ser Ala Ser Met Ala Gly Arg Gly Ser Ser Trp Ala Cys 45  
35  
Thr Leu Asn Ala Val Thr Arg Glu Gly Leu Pro Glu Xaa 60  
15 50 55

(2) INFORMATION FOR SEQ ID NO: 374:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 374:

25 Met Ser Leu Leu Asn Thr His Thr Leu Cys Phe Val Leu Phe Cys Phe 15  
1  
Thr Leu Ser Ile Asn Gln Glu Lys Leu Ala Asn His Leu Ala Phe Arg 30  
20  
Ile Leu Phe Phe Ile Val Phe Xaa 40  
35 35 40

(2) INFORMATION FOR SEQ ID NO: 375:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 44 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 375:

45 Met Cys Ser Gly Gln Ser Gln Val Trp Lys Met Ala Leu Gln Ala Leu 15  
1  
Asp Ser Glu Thr Val Val Ile Leu Pro Asp Met His Leu Ile Leu Ser 30  
50 20 25  
Leu Arg Leu Ile His Asn Ala Arg Pro Cys Leu Xaa 40  
35 40

(2) INFORMATION FOR SEQ ID NO: 376:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 203 amino acids

578

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 376:

5 Met Leu Ile Ser Glu Glu Ile Pro Phe Lys Asp Asp Pro Arg Asp 15  
1  
Glu Thr Tyr Lys Pro His Leu Glu Arg Glu Thr Pro Lys Pro Arg Arg 30  
10 Lys Ser Gly Lys Val Lys Glu Lys Glu Lys Lys Lys Glu Ile Lys Val 45  
35  
Glu Val Glu Val Glu Val Lys Glu Glu Glu Asn Glu Ile Arg Glu Asp 60  
15 50 55  
Glu Glu Pro Pro Arg Lys Arg Gly Arg Arg Arg Lys Asp Lys Ser 80  
65 70 75  
Pro Arg Leu Pro Lys Arg Arg Lys Lys Pro Pro Ile Gln Tyr Val Arg 95  
85  
Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala His Pro Arg Tyr Leu 110  
100  
Gln His His Ile Lys Tyr Gln His Leu Leu Lys Lys Lys Tyr Val Cys 125  
115 120  
Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu Gln Lys Gln Leu Leu 140  
130 135  
Arg His Ala Lys His Thr Asp Gln Arg Asp Tyr Ile Cys Glu Tyr 160  
145 150 155  
Cys Ala Arg Ala Phe Lys Ser Ser His Asn Leu Ala Val His Arg Met 175  
165 170  
Ile His Thr Gly Glu Lys His Tyr Asn Val Arg Ser Val Asp Leu Leu 190  
180  
Val Asp Lys Arg His Leu Leu Ile Gly Thr Xaa 200  
195 200

(2) INFORMATION FOR SEQ ID NO: 377:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 377:

55 Met Leu Pro Arg Arg Thr Phe Tyr Phe Thr Phe Ile Phe Phe Phe 15  
1  
Leu Ala Ser Phe Trp Gly Phe Thr Leu Arg Ala Ser Phe 25  
20 25

60

579

## (2) INFORMATION FOR SEQ ID NO: 378:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 116 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 378:

10 Met Phe Asp Ser Leu Ser Tyr Phe Lys Gly Ser Ser Leu Leu Met  
1 5 10 15  
Leu Lys Thr Tyr Leu Ser Glu Asp Val Phe Gln His Ala Val Val Leu  
20 25 30  
Tyr Leu His Asn His Ser Tyr Ala Ser Ile Gln Ser Asp Asp Leu Trp  
35 40 45  
Asp Ser Phe Asn Glu Val Thr Asn Gln Thr Leu Asp Val Lys Arg Met  
50 55 60  
Met Lys Thr Trp Thr Leu Gln Lys Gly Phe Pro Leu Val Thr Val Gln  
65 70 75 80  
Lys Lys Gly Lys Glu Leu Phe Ile Gln Gln Glu Arg Phe Phe Leu Asn  
85 90 95  
Met Lys Pro Glu Ile Gln Pro Ser Asp Thr Arg Tyr Met Pro Ser Phe  
100 105 110  
Phe Ser Cys His Leu Phe Cys Thr Leu Arg Trp Lys Tyr Phe Glu Val  
115 120 125  
Phe Tyr Asn His Lys Phe Leu Xaa  
130 135

## (2) INFORMATION FOR SEQ ID NO: 379:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 379:

45 Met Ala Trp Arg Arg Arg Glu Pro Ala Ser Gly Leu Ala Ala Cys Trp  
1 5 10 15  
Leu Trp Arg Cys Ser Pro Trp Pro Cys Ala Cys Pro Gly Pro Gly Ala  
20 25 30  
Gly Leu Ser Ser Gly Ser Arg Pro Trp  
35 40

## (2) INFORMATION FOR SEQ ID NO: 380:

## (1) SEQUENCE CHARACTERISTICS:

580

(A) LENGTH: 468 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 380:

5 Met Glu Phe Leu Lys Val Ala Arg Arg Asn Lys Arg Glu Gln Leu Glu  
1 5 10 15  
Gln Ile Gln Lys Glu Leu Ser Val Leu Glu Glu Asp Ile Lys Arg Val  
20 25 30  
Glu Glu Met Ser Gly Leu Tyr Ser Pro Val Ser Glu Asp Ser Thr Val  
35 40 45  
Pro Gln Phe Glu Ala Pro Ser Pro Ser His Ser Ser Ile Ile Asp Ser  
50 55 60  
Thr Glu Tyr Ser Gln Pro Pro Gly Phe Ser Gly Ser Ser Gln Thr Lys  
65 70 75 80  
Lys Gln Pro Trp Tyr Asn Ser Thr Leu Ala Ser Arg Arg Lys Arg Leu  
85 90 95  
Thr Ala His Phe Glu Asp Leu Glu Gln Cys Tyr Phe Ser Thr Arg Met  
100 105 110  
Ser Arg Ile Ser Asp Asp Ser Arg Thr Ala Ser Gln Leu Asp Glu Phe  
115 120 125  
Gln Glu Cys Leu Ser Lys Phe Thr Arg Tyr Asn Ser Val Arg Pro Leu  
130 135 140  
Ala Thr Leu Ser Tyr Ala Ser Asp Leu Tyr Asn Gly Ser Ser Ile Val  
145 150 155 160  
Ser Ser Ile Glu Phe Asp Arg Asp Cys Asp Tyr Phe Ala Ile Ala Gly  
165 170 175  
Val Thr Lys Lys Ile Lys Val Tyr Glu Tyr Asp Thr Val Ile Gln Asp  
180 185 190  
Ala Val Asp Ile His Tyr Pro Glu Asn Glu Met Thr Cys Asn Ser Lys  
195 200 205  
Ile Ser Cys Ile Ser Trp Ser Ser Tyr His Lys Asn Leu Leu Ala Ser  
210 215 220  
Ser Asp Tyr Glu Gly Thr Val Ile Leu Trp Asp Gly Phe Thr Gly Gln  
225 230 235 240  
Arg Ser Lys Val Tyr Gln Glu His Glu Lys Arg Cys Trp Ser Val Asp  
245 250 255  
Phe Asn Leu Met Asp Pro Lys Leu Leu Ala Ser Gly Ser Asp Asp Ala  
260 265 270  
Lys Val Lys Leu Trp Ser Thr Asn Leu Asp Asn Ser Val Ala Ser Ile  
275 280 285  
Glu Ala Lys Ala Asn Val Cys Cys Val Lys Phe Ser Pro Ser Ser Arg  
290 295 300

5 Tyr His Leu Ala Phe Gly Cys Ala Asp His Cys Val His Tyr Tyr Asp 300  
305 310 315  
Leu Arg Asn Thr Lys Lys Gln Pro Ile Met Val Phe Lys Gly His Arg Lys 320  
325 330 335

10 Ala Val Ser Tyr Ala Lys Phe Val Ser Gly Glu Ile Val Ser Ala 340 345 350

Ser Thr Asp Ser Gln Leu Lys Leu Trp Asn Val Gly Pro Tyr Cys 355 360 365

15 Leu Arg Ser Phe Lys Gly His Ile Asn Glu Lys Asn Phe Val Gly Leu  
170 375 380

Ala Ser Asn Gly Asp Tyr Ile Ala Cys Gly Ser Glu Asn Asn Ser Leu	
185	390
	395
	400

Tyr Leu Tyr Tyr Lys Gly Leu Ser Lys Thr Leu Leu Thr Phe Lys Phe  
405 410 415

25 Asp Thr Val Lys Ser Val Leu Asp Lys Asp Arg Lys Glu Asp Asp Thr  
420 430

Asn Glu Phe Val Ser Ala Val Cys Trp Arg Ala Leu Pro Asp Gly Glu  
435 440 445

30 Ser Asn Val Leu Ile Ala Ala Asn Ser Gln Gly Thr Ile Lys Val Leu  
450 . 455 460

**Glu Leu Val Xaa**  
**465**

(2) INFORMATION FOR SEQ ID NO: 381:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

45 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 361:  
Met Arg Lys Glu Asp Gly Phe Trp Phe Phe Phe Leu Phe Phe  
1 5 10 15

Val Val Gly Ser Lys Phe Val Asn Gly Asn Lys Leu Val  
50 20 25

(2) INFORMATION FOR SEQ ID NO: 382:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

Met Pro Leu Ala Pro Tyr Cys Asp Leu Val Ala Leu Ser Phe Ala  
1 5 10 15

5 Leu Val Glu Ser Pro Val Asp Ser Ser Asp Phe Thr  
20 25

10 (2) INFORMATION FOR SEQ ID NO: 383:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 138 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 383:

20 Met Asn Ser Leu Val Ser Trp Cln Leu Leu Phe Leu Cys Ala Thr  
1 5 10 15  
His Phe Gly Cln Pro Leu Glu Lys Val Ala Ser Val Gly Asn Ser Arg  
20 25 30

Pro Thr Gly Gln Gln Leu Glu Ser Leu Gly Leu Ala Pro Gly Glu  
25 35 40 45

Gln Ser Leu Pro Cys Thr Glu Arg Lys Pro Ala Ala Thr Ala Arg Leu  
50 55 60

30 Ser Arg Arg Gly Thr Ser Leu Ser Pro Pro Glu Ser Ser Gly Ser  
65 70 75 80

Pro Gln Gln Pro Gly Leu Ser Ala Pro His Ser Arg Gln Ile Pro Ala  
85 90 95

Pro Gln Gly Ala Val Leu Val Gln Arg Glu Lys Asp Leu Pro Asn Tyr  
100 105 110

40 Asn Trp Asn Ser Phe Gly Leu Arg Phe Gly Lys Arg Glu Ala Ala Pro  
115 120 125

Gly Asn His Gly Arg Ser Ala Gly Arg Gly  
130 135

50

(2) INFORMATION FOR SEQ ID NO: 384:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(iii) STRUCTURE INFORMATION: SEQ ID NO: 384:

55 Met Ser Cys Phe Ile Asp Ser Xaa Asp Ser Lys Ile Leu His Leu Leu 15  
1 5 10  
Val Val Ser Phe Ile Cys Xaa Leu Phe Leu Ile Leu Thr His Gly 10  
20 25



11e Leu Ile Leu Arg Xaa Phe Ser Val Xaa Xaa His Ser Leu Lys  
35 40 45  
Asn Asn Leu Glu Glu Tyr Leu Ile Leu Met Asn Lys Ala Leu Leu Thr  
5 50 55 60  
Arg Glu Asp Phe Phe Val Leu Pro Xaa Ala  
65 70  
(2) INFORMATION FOR SEQ ID NO: 385:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 521 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 385:  
Met Ser Ala Gly Glu Val Glu Arg Leu Val Ser Glu Leu Ser Gly Gly  
1 5 10 15  
Thr Gly Gly Asp Glu Glu Glu Glu Tyr Leu Tyr Gly Asp Glu Asn Glu  
20 25 30  
Val Glu Arg Pro Glu Glu Glu Asn Ala Ser Ala Asn Pro Ser Gly  
35 40 45  
Ile Glu Asp Glu Thr Ala Glu Asn Gly Val Pro Lys Pro Lys Val Thr  
50 55 60  
Glu Thr Glu Asp Asp Ser Asp Ser Asp Asp Asp Glu Asp Asp  
65 70 75 80  
Val His Val Thr Ile Gly Asp Ile Lys Thr Gly Ala Pro Gln Tyr Gly  
85 90 95  
Ser Tyr Gly Thr Ala Pro Val Asn Leu Asn Ile Lys Thr Gly Gly Arg  
100 105 110  
Val Tyr Gly Thr Thr Gly Thr Lys Val Lys Gly Val Asp Leu Asp Ala  
115 120 125  
Pro Gly Ser Ile Asn Gly Val Pro Leu Leu Glu Val Asp Leu Asp Ser  
130 135 140  
Phe Glu Asp Lys Pro Tyr Arg Lys Pro Gly Ala Asp Leu Ser Asp Tyr  
145 150 155 160  
Phe Asn Tyr Gly Phe Asn Glu Asp Thr Tyr Lys Ala Tyr Cys Glu Lys  
165 170 175  
Gln Lys Arg Ile Arg Met Gly Leu Glu Val Ile Pro Val Thr Ser Thr  
180 185 190  
Thr Asn Lys Ile Thr Val Gln Gln Gly Arg Thr Gly Asn Ser Glu Lys  
195 200 205  
Glu Thr Ala Leu Pro Ser Thr Lys Ala Glu Phe Thr Ser Pro Pro Ser  
210 215 220

Leu Phe Lys Thr Gly Leu Pro Pro Ser Arg Arg Leu Pro Gly Ala Ile  
225 230 235 240  
Asp Val Ile Gly Gln Thr Ile Thr Ile Ser Arg Val Glu Gly Arg Arg  
245 250 255  
Arg Ala Asn Glu Asn Ser Asn Ile Gln Val Leu Ser Glu Arg Ser Ala  
260 265 270  
Thr Glu Val Asp Asn Asn Phe Ser Lys Pro Pro Phe Phe Pro  
275 280 285  
Gly Ala Pro Thr His Leu Pro Pro Pro Pro Phe Leu Pro Pro  
290 295 300  
Pro Thr Val Ser Thr Ala Pro Pro Leu Ile Pro Pro Gly Phe Pro  
305 310 315 320  
Pro Pro Pro Gly Ala Pro Pro Ser Ser Leu Ile Pro Thr Ile Glu Ser  
325 330 335  
Gly His Ser Ser Gly Tyr Asp Ser Arg Ser Ala Arg Ala Phe Pro Tyr  
340 345 350  
Gly Asn Val Ala Phe Pro His Leu Pro Gly Ser Ala Pro Ser Thr Pro  
355 360 365  
Ser Leu Val Asp Thr Ser Lys Gln Tyr Asp Tyr Tyr Ala Arg Arg Glu  
370 375 380  
Lys Asp Arg Asp Arg Glu Arg Asp Arg Asp Arg Glu Asp Asp  
385 390 395 400  
Arg Asp Arg Glu Arg Glu Arg Thr Arg Glu Arg Glu Arg Asp  
405 410 415  
His Ser Pro Thr Pro Ser Val Phe Asn Ser Asp Glu Glu Arg Tyr Arg  
420 425 430  
Tyr Arg Glu Tyr Ala Glu Arg Gly Tyr Glu Arg His Arg Ala Ser Arg  
435 440 445  
Glu Lys Glu Glu Arg His Arg Glu Arg Arg His Arg Glu Lys Glu Glu  
450 455 460  
Thr Arg His Lys Ser Ser Arg Ser Asn Ser Arg Arg His Glu Ser  
465 470 475 480  
Glu Glu Gly Asp Ser His Arg Arg His Lys His Lys Ser Lys Arg  
485 490 495  
Ser Lys Glu Gly Lys Glu Ala Gly Ser Glu Pro Ala Pro Glu Gln Glu  
500 505 510  
Ser Thr Glu Ala Thr Pro Ala Glu Xaa  
515 520

## (2) INFORMATION FOR SEQ ID NO: 386:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 137 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 386:

Met Asn Ser Arg Gly Ile Trp Leu Ala Tyr Ile Ile Leu Val Gly Leu 15  
 1 5 10  
 Leu His Met Val Leu Leu Ser Ile Pro Phe Phe Ser Ile Pro Val Val 30  
 20 25  
 Trp Thr Leu Thr Asn Val Ile His Asn Leu Ala Thr Tyr Val Phe Leu 45  
 35 40  
 His Thr Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala 60  
 50 55  
 Arg Leu Leu Thr His Trp Glu Gln Met Asp Tyr Gly Leu Gln Phe Thr 80  
 65 70 75  
 Ser Ser Arg Lys Phe Leu Ser Ile Ser Pro Ile Val Leu Tyr Leu Leu 95  
 85 90  
 Ala Ser Phe Tyr Thr Lys Tyr Asp Ala Ala His Phe Leu Ile Asn Thr 110  
 100 105  
 Ala Ser Leu Leu Ser Val Leu Leu Pro Lys Leu Pro Gln Phe His Gly 125  
 115 120  
 Val Arg Val Phe Gly Ile Asn Lys Tyr 135  
 130

35

## (2) INFORMATION FOR SEQ ID NO: 387:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 186 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 387:

Met Ala Ala Gln Lys Asp Gln Gln Lys Asp Ala Glu Ala Glu Gly Leu 15  
 1 5 10  
 Ser Gly Thr Thr Leu Leu Pro Lys Leu Ile Pro Ser Gly Ala Gly Arg 30  
 20 25  
 Glu Trp Leu Glu Arg Arg Ala Thr Ile Arg Pro Trp Ser Thr Phe 45  
 35 40  
 Val Asp Gln Gln Arg Phe Ser Arg Pro Arg Asn Leu Gly Glu Leu Cys 60  
 50 55  
 Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe 80  
 65 70 75

60

Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met Leu 95  
 85 90  
 Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Tyr Ile Leu Tyr Leu 110  
 100 105  
 Arg Thr Leu Glu Ser Lys Leu Val Leu Phe Gly Arg Glu Val Ser Pro 125  
 115 120  
 Ala His Gln Tyr Ala Leu Ala Gly Ile Ser Phe Pro Phe Thr Trp 140  
 130 135  
 Leu Ala Gly Ala Gly Ser Ala Val Phe Trp Val Leu Gly Ala Thr Leu 160  
 145 150 155  
 Val Val Ile Gly Ser His Ala Ala Phe His Gln Ile Glu Ala Val Asp 175  
 165 170  
 Gly Glu Glu Leu Gln Met Glu Pro Val Xaa 185  
 180

20

## (2) INFORMATION FOR SEQ ID NO: 388:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 388:

Met  
 1

35

## (2) INFORMATION FOR SEQ ID NO: 389:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 299 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 389:

Met Leu Ser Ile Phe Tyr Phe Ala Ile Pro Val Gly Ser Gly Leu Gly 15  
 1 5 10  
 Tyr Ile Ala Gly Ser Lys Val Lys Asp Met Ala Gly Asp Trp His Trp 30  
 20 25  
 Ala Leu Arg Val Thr Pro Gly Leu Gly Val Val Ala Val Leu Leu Leu 45  
 35 40  
 Phe Leu Val Val Arg Glu Pro Pro Arg Gly Ala Val Glu Arg His Ser 60  
 50 55  
 Asp Leu Pro Pro Leu Asn Pro Thr Ser Trp Ala Asp Leu Arg Ala 80  
 65 70 75  
 Leu Ala Arg Asn Pro Phe Val Leu Ser Ser Leu Gly Phe Thr Ala 90

85 90 95  
Val Ala Phe Val Thr Gly Ser Leu Ala Leu Trp Ala Pro Ala Phe Leu  
100 105 110  
5 Leu Arg Ser Arg Val Val Leu Gly Glu Thr Pro Cys Leu Pro Gly  
115 120 125  
Asp Ser Cys Ser Ser Asp Ser Leu Ile Phe Gly Leu Ile Thr Cys  
130 135 140  
10 Leu Thr Gly Val Leu Gly Val Gly Leu Gly Val Glu Ile Ser Arg Arg  
145 150 155 160  
Leu Arg His Ser Asn Pro Arg Ala Asp Pro Leu Val Cys Ala Thr Gly  
165 170 175  
15 Leu Leu Gly Ser Ala Pro Phe Leu Phe Leu Ser Leu Ala Cys Ala Arg  
180 185 190  
20 Gly Ser Ile Val Ala Thr Tyr Ile Phe Ile Phe Ile Gly Glu Thr Leu  
195 200 205  
25 Leu Ser Met Asn Trp Ala Ile Val Ala Asp Ile Leu Leu Tyr Val Val  
210 215 220  
Ile Pro Thr Arg Arg Ser Thr Ala Glu Ala Phe Glu Ile Val Leu Ser  
225 230 235 240  
30 His Leu Leu Gly Asp Ala Gly Ser Pro Tyr Leu Ile Gly Leu Ile Ser  
245 250 255  
Asp Arg Leu Arg Arg Asn Trp Pro Pro Ser Phe Leu Ser Glu Phe Arg  
260 265 270  
35 Ala Leu Glu Phe Ser Leu Met Leu Cys Ala Phe Val Gly Ala Leu Gly  
275 280 285  
40 Gly Ala Leu Pro Gly His Arg His Leu His Xaa  
290 295

(2) INFORMATION FOR SEQ ID NO: 390:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 49 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 390:  
Met Gly Pro Glu Gly Trp Val Arg Pro Leu Lys Thr Ala Pro Lys Leu  
1 5 10 15  
Gly Glu Ala Ile Arg Leu Ile Leu Phe Leu Asn Phe Val Lys Glu Cys  
20 25 30  
Ile Ala Ser Val Asn Leu Cys Ile Leu Arg Leu Asn Ile Thr Pro Leu  
35 40 45 60

Leu  
5 (2) INFORMATION FOR SEQ ID NO: 391:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 61 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 391:  
Met Tyr Val Asn Tyr Gly Thr Arg Asn Tyr Ser Thr Glu Gly Pro Ala  
1 5 10 15  
Ala Leu Leu Asp Glu Ala Lys Leu Ser Leu Leu Val Trp Val Leu Cys  
20 25 30  
20 Phe Val Leu Leu Phe Val Cys Phe Cys Gly Leu Ser Tyr Val Val Ile  
35 40 45  
25 Ala Glu Val Pro Val Gly Leu Leu Cys Ile Thr Glu Xaa  
50 55 60

(2) INFORMATION FOR SEQ ID NO: 392:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 79 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 392:  
Met Leu Trp Phe Ala Asn Phe Phe Thr Tyr Leu Phe Leu Ser Glu Ser  
1 5 10 15  
Val Ala Phe Val His Ile Ser His Ile Gly Val Arg Glu Val Asn Thr  
20 25 30  
Asn Cys Tyr Phe Ser Arg Lys Ser Tyr Cys Tyr Gly Ile Leu Asn Pro  
35 40 45  
45 Ile Asn Cys Ile Lys Gly Lys Lys Lys Lys Lys Lys Lys Lys  
50 55 60  
Lys Lys Lys Lys Ile Pro Ala Gly Arg Xaa Leu Phe Pro Phe Gly  
65 70 75 80

(2) INFORMATION FOR SEQ ID NO: 393:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 393:  
Met Tyr Val Asn Tyr Gly Thr Arg Asn Tyr Ser Thr Glu Gly Pro Ala  
1 5 10 15  
Ala Leu Leu Asp Glu Ala Lys Leu Ser Leu Leu Val Trp Val Leu Cys  
20 25 30  
20 Phe Val Leu Leu Phe Val Cys Phe Cys Gly Leu Ser Tyr Val Val Ile  
35 40 45  
25 Ala Glu Val Pro Val Gly Leu Leu Cys Ile Thr Glu Xaa  
50 55 60

589

Met Pro Gly Ala Phe Ser Glu Thr Val Ile Asn Asp Leu Leu Ser Leu  
1 5 10 15

Phe Leu Val Leu Pro Ala Glu Ser Tyr Ser Thr Leu Ser Gly Val  
20 25 30

Tyr Arg Asn Ala  
35

10

(2) INFORMATION FOR SEQ ID NO: 394:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 180 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 394:

Met Ala Gln Ser Arg Asp Gly Gly Asn Pro Phe Ala Glu Pro Ser Glu  
1 5 10 15

Leu Asp Asn Pro Phe Gln Asp Pro Ala Val Ile Gln His Arg Pro Ser  
20 25 30

Arg Gln Tyr Ala Thr Leu Asp Val Tyr Asn Pro Phe Glu Thr Arg Glu  
35 40 45

Pro Pro Pro Ala Tyr Glu Pro Pro Ala Pro Ala Pro Leu Pro Pro Pro  
50 55 60

Ser Ala Pro Ser Leu Gln Pro Ser Arg Lys Leu Ser Pro Thr Glu Pro  
65 70 75 80

Lys Asn Tyr Gly Ser Tyr Ser Thr Gln Ala Ser Ala Ala Ala Thr  
85 90 95

Ala Glu Leu Leu Lys Lys Gln Glu Glu Leu Asn Arg Lys Ala Glu Glu  
100 105 110

Leu Asp Arg Arg Ser Glu Ser Cys Ser Met Leu Pro Trp Xaa Ala Gln  
115 120 125

Leu Leu Asp Arg Thr Ile Gly Pro Leu Tyr Leu Phe Val Gln Phe  
130 135 140

Ser Pro Ala Phe Ser Arg Thr Ser Pro Trp Arg Ser Pro Lys Asn Phe  
145 150 155 160

Arg Arg Leu Tyr Pro Pro Cys Thr Thr Ser Gly Cys Ala Ala Arg Trp  
165 170 175

Xaa Phe Ser Xaa  
180

55

(2) INFORMATION FOR SEQ ID NO: 395:

(i) SEQUENCE CHARACTERISTICS:

60

590

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 395:

Met Pro Thr Pro Cys Thr Ser Leu Pro Ser Cys Cys Gln His Arg Ser  
1 5 10 15

Ile Thr Met Thr Leu  
20

10

(2) INFORMATION FOR SEQ ID NO: 396:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 396:

Met Pro Leu Phe Ile Pro Leu Ile Phe Leu Ser Leu Leu His Cys  
1 5 10 15

Gln Ser Lys His Pro Ile Gln Met Ser Leu Cys Met Cys Val Asn Ile  
20 25 30

Ser Leu Val Trp Ser Pro Val Arg Trp Ile Phe Gly Ser Lys Gly Leu  
35 40 45

Phe Ser Val His Leu Gln Ser Ser Gln Arg Pro Ser  
50 55 60

35

(2) INFORMATION FOR SEQ ID NO: 397:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 152 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 397:

Met Ala Gly Pro Arg Pro Xaa Trp Arg Asp Gln Leu Leu Phe Met Ser  
1 5 10 15

Ile Ile Val Leu Val Ile Val Ile Cys Leu Met Leu Tyr Ala Leu  
20 25 30

Leu Trp Glu Ala Gly Asn Leu Thr Asp Leu Pro Asn Leu Arg Ile Gly  
35 40 45

Phe Tyr Asn Phe Cys Leu Trp Asn Glu Asp Thr Ser Thr Leu Gln Cys  
50 55 60

His Gln Phe Pro Glu Leu Glu Ala Leu Gly Val Pro Arg Val Gly Leu  
65 70 75 80

Gly Leu Ala Arg Leu Gly Val Tyr Gly Ser Leu Val Leu Thr Leu Phe  
85 90 95

Ala Pro Gln Pro Leu Leu Leu Ala Gln Cys Asn Xaa Asp Glu Arg Ala  
100 105 110  
5 Trp Arg Leu Ala Val Gly Phe Leu Ala Val Ser Ser Val Leu Leu Ala  
115 120 125  
10 Gly Gly Leu Gly Leu Phe Leu Ser Tyr Val Trp Asn Gly Ser Xaa Ser  
130 135 140  
Pro Ser Arg Gly Leu Gly Phe Xaa  
145 150  
15 (2) INFORMATION FOR SEQ ID NO: 398:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 480 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 398:  
20 Met Ser Asp Gly Phe Asp Arg Ala Pro Gly Ala Gly Arg Gly Arg Xaa  
1 5 10 15  
Arg Gly Leu Gly Arg Gly Gly Gly Pro Xaa Gly Gly Phe Pro  
20 25 30  
30 Xaa Gly Xaa Xaa Pro Ala Glu Arg Xaa Arg His Gln Pro Pro Gln Pro  
35 40 45  
Lys Ala Pro Gly Phe Leu Gln Pro Xaa Pro Leu Arg Gln Pro Arg Thr  
50 55 60  
35 Thr Pro Pro Gly Ala Gln Cys Glu Val Pro Ala Ser Pro Gln Arg  
65 70 75 80  
40 Pro Ser Arg Pro Gly Ala Leu Pro Glu Gln Thr Arg Pro Leu Arg Ala  
85 90 95  
Pro Pro Ser Ser Gln Asp Lys Ile Pro Gln Gln Asn Ser Glu Ser Ala  
100 105 110  
45 Met Ala Lys Pro Gln Val Val Ala Pro Val Leu Met Ser Lys Leu  
115 120 125  
Ser Val Asn Ala Pro Glu Phe Tyr Pro Ser Gly Tyr Ser Ser Ser Tyr  
130 135 140  
50 Thr Glu Ser Tyr Glu Asp Gly Cys Glu Asp Tyr Pro Thr Leu Ser Glu  
145 150 155 160  
Tyr Val Gln Asp Phe Leu Asn His Leu Thr Glu Gln Pro Gly Ser Phe  
165 170 175  
Glu Thr Glu Ile Glu Gln Phe Ala Glu Thr Leu Asn Gly Cys Val Thr  
180 185 190  
60 Thr Asp Asp Ala Leu Gln Glu Leu Val Glu Leu Ile Tyr Gln Gln Ala

195 200 205  
5 Thr Ser Ile Pro Asn Phe Ser Tyr Met Gly Ala Arg Leu Cys Asn Tyr  
210 215 220  
Leu Ser His His Leu Thr Ile Ser Pro Gln Ser Gly Asn Phe Arg Gln  
225 230 235 240  
10 Leu Leu Leu Gln Arg Cys Arg Thr Glu Tyr Glu Val Lys Asp Gln Ala  
245 250 255  
Ala Lys Gly Asp Glu Val Thr Arg Lys Arg Phe His Ala Phe Val Leu  
260 265 270  
15 Phe Leu Gly Glu Leu Tyr Leu Asn Leu Ile Lys Gly Thr Asn Gly  
275 280 285  
Gln Val Thr Arg Ala Asp Ile Leu Gln Val Gly Leu Arg Glu Leu Leu  
290 295 300  
20 Asn Ala Leu Phe Ser Asn Pro Met Asp Asp Asn Leu Ile Cys Ala Val  
305 310 315 320  
Lys Leu Leu Lys Leu Thr Gly Ser Val Leu Glu Asp Ala Trp Lys Glu  
325 330 335  
25 Lys Gly Lys Met Asp Met Glu Glu Ile Ile Gln Arg Ile Glu Asn Val  
340 345 350  
30 Val Leu Asp Ala Asn Cys Ser Arg Asp Val Lys Gln Met Leu Lys  
355 360 365  
Leu Val Glu Leu Arg Ser Ser Asn Trp Gly Arg Val His Ala Thr Ser  
370 375 380  
35 Thr Tyr Arg Glu Ala Thr Glu Asn Asp Pro Asn Tyr Phe Met Asn  
385 390 395 400  
Glu Pro Thr Phe Tyr Thr Ser Asp Gly Val Pro Phe Thr Ala Ala Asp  
405 410 415  
40 Pro Asp Tyr Gln Glu Lys Tyr Gln Glu Leu Leu Glu Arg Glu Asp Phe  
420 425 430  
Phe Pro Asp Tyr Glu Glu Asn Gly Thr Asp Leu Ser Gly Ala Gly Asp  
435 440 445  
50 Pro Tyr Leu Asp Asp Ile Asp Asp Glu Met Asp Pro Glu Ile Glu Glu  
450 455 460  
Ala Tyr Glu Lys Phe Cys Leu Glu Ser Glu Arg Lys Arg Lys Gln Xaa  
465 470 475 480  
55 (2) INFORMATION FOR SEQ ID NO: 399:  
60

WO 98/39448

594

Ile Thr Leu Met Gly Pro Gln Trp Trp Leu Lys Thr Val Ile Glu Gln  
290 295 300  
Val Tyr Ala Asn Gly Ile Arg Asn Ile Asp Leu His Tyr Ile Val Arg  
305 310 315  
Lys Leu Ala Ala Pro Val Ile Ser Val Leu Leu Ser Leu Cys Val  
320 325 330 335  
10 Pro Tyr Val Ile Ala Ser Gly Val Val Pro Leu Leu Gly Val Thr Ala  
340 345 350  
Glu Met Gln Asn Leu Val His Arg Arg Ile Tyr Pro Phe Leu Leu Met  
355 360 365  
15 Val Val Val Leu Met Ala Ile Leu Ser Phe Gln Val Arg Gln Phe Lys  
370 375 380  
Arg Leu Tyr Glu His Ile Lys Asn Asp Lys Tyr Leu Val Gly Gln Arg  
385 390 395 400  
Leu Val Asn Tyr Glu Arg Lys Ser Gly Lys Gln Gly Ser Ser Pro Pro  
405 410 415  
25 Pro Pro Gln Ser Ser Gln Glu  
420

(1) INFORMATION FOR SEQ ID NO: 400:

(A) LENGTH: 78 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 400:

Met Leu Arg Leu Asp Ile Ile Asn Ser Leu Val Thr Thr Val Phe Met  
1 5 10 15  
Leu Ile Val Ser Val Leu Ala Leu Ile Pro Glu Thr Thr Thr Leu Thr  
20 25 30  
Val Gly Gly Val Phe Ala Leu Val Thr Ala Val Cys Cys Leu Ala  
35 40 45  
Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro Ser Gly Pro  
50 55 60  
50 Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu Xaa  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 401:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

PCT/US98/04493

WO 98/39448

593

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 423 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 399:

Met Glu Pro Lys Thr Ile Thr Asp Ala Leu Ala Ser Ser Ile Ile Lys  
1 5 10 15  
10 Ser Val Leu Pro Asn Phe Leu Pro Tyr Asn Val Met Leu Tyr Ser Asp  
20 25 30  
Ala Pro Val Ser Glu Leu Ser Leu Glu Leu Leu Leu Gln Val Val  
35 40 45  
15 Leu Pro Ala Leu Leu Glu Gln Gly His Thr Arg Gln Trp Leu Lys Gly  
50 55 60  
Leu Val Arg Ala Trp Thr Val Thr Ala Gly Tyr Leu Leu Asp Leu His  
65 70 75 80  
20 Ser Tyr Leu Leu Gly Asp Gln Glu Glu Asn Ser Ala Asn Gln  
85 90 95  
25 Gln Val Asn Asn Asn Gln His Ala Arg Asn Asn Ala Ile Pro Val  
100 105 110  
Val Gly Gly Leu His Ala Ala His Gln Ala Ile Leu Gln Gln Gly  
115 120 125  
30 Gly Pro Val Gly Phe Gln Xaa Tyr Arg Pro Leu Asn Phe Pro Leu  
130 135 140  
Arg Ile Phe Leu Leu Ile Val Phe Met Cys Ile Thr Leu Leu Ile Ala  
145 150 155 160  
35 Ser Leu Ile Cys Leu Thr Leu Pro Val Phe Ala Gly Arg Trp Leu Met  
165 170 175  
40 Ser Phe Trp Thr Gly Thr Ala Lys Ile His Glu Leu Tyr Thr Ala Ala  
180 185 190  
Cys Gly Leu Tyr Val Cys Trp Leu Thr Ile Arg Ala Val Thr Val Met  
195 200 205  
45 Val Ala Trp Met Pro Gln Gly Arg Val Ile Phe Gln Lys Val Lys  
210 215 220  
Glu Trp Ser Leu Met Ile Met Lys Thr Leu Ile Val Ala Val Leu Leu  
225 230 235 240  
50 Ala Gly Val Val Pro Leu Leu Leu Gly Leu Leu Phe Glu Leu Val Ile  
245 250 255  
55 Val Ala Pro Leu Arg Val Pro Leu Asp Gln Thr Pro Leu Phe Tyr Pro  
260 265 270  
Trp Gln Asp Trp Ala Leu Gly Val Leu His Ala Lys Ile Ile Ala Ala  
275 280 285  
60

595

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 401:

Met Leu Lys Gln Val Met Phe Val Phe Ser Gly Met Gly Pro Arg Ser  
1 5 10 15  
His Cys Trp Gly Leu Pro Leu His Val Ala Pro Leu Cys Arg Gly His  
20 25 30  
Gln Ala Asp Ser Ser His Leu Leu Pro Leu Lys His Gln Gly Ala Trp  
35 40 45  
Asn Arg Asn Leu Ala Asn Gln Arg His Phe Phe Cys Pro Ser Ile Phe  
50 55 60  
His Thr Cys Pro Thr Val Leu Leu Phe Xaa  
65 70

20 (2) INFORMATION FOR SEQ ID NO: 402:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 402:

Ala Arg Thr Ile Leu Val Leu Tyr Leu Ser Leu Gln Arg Leu Gln Asn  
1 5 10 15  
Leu Ala Tyr His  
20

35 (2) INFORMATION FOR SEQ ID NO: 403:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 87 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 403:

Met Pro Leu Pro Ser Val Pro Ile Leu Gly Ile Phe Ser Phe Leu Ile  
1 5 10 15  
Pro Ser Ser Gln Gly Val Ser Tyr Thr Lys Leu Pro Ile Ser Ser Pro  
20 25 30

Gln Tyr Ser Pro Phe Val Asn Asp His Phe Ser Phe Leu Asn Pro Phe  
35 40 45

Pro Val Gln Ile His Thr Gly Phe Ala Arg Val Gly Ser Tyr Met Gln  
50 55 60

Met Pro Leu Val His Leu Cys Leu Leu Gln Thr Ser Leu Met Lys Asn  
65 70 75 80

Ser Gly Val Gln Gln Gly Ser  
85

596

(2) INFORMATION FOR SEQ ID NO: 404:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 404:

Met Asn Ala Ala Met Val His Ile Asn Arg Ala Leu Lys Leu Ile Ile  
1 5 10 15

Arg Leu Phe Leu Val Glu Asp Leu Val Asp Ser Leu Lys Leu Ala Val  
20 25 30

Phe Met Trp Leu Met Thr Tyr Val Gly Ala Val Phe Asn Gly Ile Thr  
35 40 45

Leu Leu Ile Leu Ala Glu Leu Leu Ile Phe Ser Val Pro Ile Val Tyr  
50 55 60

Glu Lys Tyr Lys Thr Gln Ile Asp His Tyr Val Gly Ile Ala Arg Asp  
65 70 75 80

Gln Thr Lys Ser Ile Val Glu Lys Ile Pro Ser Lys  
85 90

30 (2) INFORMATION FOR SEQ ID NO: 405:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 405:

Met Ala Cys Ser Cys Leu Met Ile Gln Ser Phe Ser Thr Ser Ala Leu  
1 5 10 15

Val Leu Phe Tyr Gly  
20

45 (2) INFORMATION FOR SEQ ID NO: 406:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 174 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 406:

Met Glu Glu Gly Gly Asn Leu Gly Gly Leu Ile Lys Met Val His Leu  
1 5 10 15

Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp Val Thr Phe Val  
20 25 30

Ser Gly Phe Pro Ala Phe Pro Lys Pro Ser Pro Thr Tyr Leu Arg Thr  
35 40 45

5 Ser Ala Glu Gln Thr Leu Pro Leu Leu Pro His Leu His Gly Leu  
50 55 60

Cys Leu His Gln Pro Leu His Leu Gly Phe Thr Ala Cys Leu Gly Ser  
65 70 75 80

10 Ala His Ile Leu Gly Gln Pro Ala Leu Pro Ala Val Pro Glu Pro  
85 90 95

15 Tyr Ala Gly His Cys Gln Arg Pro Leu Ala Gly Thr Pro His His Ser  
100 105 110

Cys His Val Gly Pro Ala Asn Arg Gly Arg Arg Ser Glu Ala Trp Val  
115 120 125

20 Gly Arg Tyr Gln Ala Ala Asn Arg Phe Pro Ile Leu Asn Ala Xaa Cys  
130 135 140

Glu Arg Arg Thr Pro Ser Thr Val Leu Ser Ala Arg Ile Ser Ser Ala  
145 150 155 160

25 Thr Met Gly Cys Pro Leu Phe Ala Ile Trp Ala Ala Ser Xaa  
165 170

30 (2) INFORMATION FOR SEQ ID NO: 407:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 407:

Met Ala Phe Ile Leu Leu Phe Tyr Cys Leu Met Thr Phe Leu Ser Leu  
1 5 10 15

Glu Gln Asn Ser Ala Thr Val Glu Pro Ser Ser His Glu Ile Leu His  
20 25 30

45 Leu Leu Gln Asn Cys Phe Glu Leu Leu Arg Thr Ser Thr Ser Gln Cys  
35 40 45

Thr Glu Gly Ile Pro Cys Gln Arg Tyr Gln Asn Gly Leu His Ile Xaa  
50 55 60

55 (2) INFORMATION FOR SEQ ID NO: 408:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 280 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 408:

Met Glu Ala Val Val Asn Leu Tyr Gln Glu Val Met Lys His Ala Asp  
1 5 10 15

Pro Arg Ile Gln Gly Tyr Pro Leu Met Gly Ser Pro Leu Leu Met Thr  
20 25 30

10 Ser Ile Leu Leu Thr Tyr Val Tyr Phe Val Leu Ser Leu Gly Pro Arg  
35 40 45

Ile Met Ala Asn Arg Lys Pro Phe Gln Leu Arg Gly Phe Met Ile Val  
50 55 60

Tyr Asn Phe Ser Leu Val Ala Leu Ser Leu Tyr Ile Val Tyr Glu Phe  
65 70 75 80

Leu Met Ser Gly Trp Leu Ser Thr Tyr Thr Trp Arg Cys Asp Pro Val  
85 90 95

20 Asp Tyr Ser Asn Ser Pro Glu Ala Leu Arg Met Val Arg Val Ala Trp  
100 105 110

25 Leu Phe Leu Phe Ser Lys Phe Ile Glu Leu Met Asp Thr Val Ile Phe  
115 120 125

Ile Leu Arg Lys Lys Asp Gly Gln Val Thr Phe Leu His Val Phe His  
130 135 140

30 His Ser Val Leu Pro Trp Ser Trp Trp Trp Gly Val Lys Ile Ala Pro  
145 150 155 160

Gly Gly Met Gly Ser Phe His Ala Met Ile Asn Ser Ser Val His Val  
165 170 175

Ile Met Tyr Leu Tyr Tyr Gly Leu Ser Ala Phe Gly Pro Val Ala Gln  
180 185 190

40 Pro Tyr Leu Trp Trp Lys Lys His Met Thr Ala Ile Gln Leu Ile Gln  
195 200 205

Phe Val Leu Val Ser Leu His Ile Ser Gln Tyr Tyr Phe Met Ser Ser  
210 215 220

45 Cys Asn Tyr Gln Tyr Pro Val Ile Ile His Leu Ile Trp Met Tyr Gly  
225 230 235 240

Thr Ile Phe Phe Met Leu Phe Ser Asn Phe Trp Tyr His Ser Tyr Thr  
245 250 255

Lys Gly Lys Arg Leu Pro Arg Ala Leu Gln Gln Asn Gly Ala Pro Gly  
260 265 270

55 Ile Ala Lys Val Lys Ala Asn Xaa  
275 280

60 (2) INFORMATION FOR SEQ ID NO: 409:



## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 284 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 409:

Met Xaa Leu Trp Pro Gln Thr Cys Ser Gly Lys Phe Asp Gly Thr Leu  
1 5 10 15  
Ala Phe Ser Ile His Xaa Leu Ala Val Ile Leu Gly Asp Gln Leu Thr  
20 25 30  
Ala Ala Asp Leu Val Pro Ile Phe Asn Gly Phe Leu Lys Asp Leu Asp  
35 40 45  
Glu Val Arg Ile Gly Val Leu Lys His Leu His Asp Phe Leu Lys Leu  
50 55 60  
Leu His Ile Asp Lys Arg Arg Glu Tyr Leu Tyr Gln Leu Gln Glu Phe  
65 70 75 80  
Leu Val Thr Asp Asn Ser Arg Asn Trp Arg Phe Arg Ala Glu Leu Ala  
85 90 95  
Glu Gln Leu Ile Leu Leu Leu Glu Leu Tyr Ser Pro Arg Asp Val Tyr  
100 105 110  
Asp Tyr Leu Arg Pro Ile Ala Leu Asn Leu Cys Ala Asp Lys Val Ser  
115 120 125  
Ser Val Arg Trp Ile Ser Tyr Lys Leu Val Ser Glu Met Val Lys Lys  
130 135 140  
Leu His Ala Ala Thr Pro Pro Thr Phe Gly Val Asp Leu Ile Asn Glu  
145 150 155 160  
Leu Val Glu Asn Phe Gly Arg Cys Pro Lys Trp Ser Gly Arg Gln Ala  
165 170 175  
Phe Val Phe Val Cys Gln Thr Val Ile Glu Asp Asp Cys Leu Pro Met  
180 185 190  
Asp Gln Phe Ala Val His Leu Met Pro His Leu Leu Thr Leu Ala Asn  
195 200 205  
Asp Arg Val Pro Asn Val Arg Val Leu Leu Ala Lys Thr Leu Arg Gln  
210 215 220  
Thr Leu Leu Glu Lys Asp Tyr Phe Leu Ala Ser Ala Ser Cys His Gln  
225 230 235 240  
Glu Ala Val Glu Gln Thr Thr Ile Met Ala Leu Gln Met Asp Arg Asp Ser  
245 250 255  
Asp Val Lys Tyr Phe Ala Ser Ile His Pro Ala Ser Thr Lys Ile Ser  
260 265 270  
Glu Asp Ala Met Ser Thr Ala Ser Ser Thr Tyr Xaa  
275 280

## (2) INFORMATION FOR SEQ ID NO: 410:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 187 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 410:

Met Leu Phe Leu Phe Phe Val Ile Ile Phe Leu Phe Val Phe Leu Ile  
1 5 10 15  
Leu Ile Ile Gln Phe Ser Lys Pro Leu Thr Asn Pro His Pro Pro Ala  
20 25 30  
Gly Xaa Ser Asp Arg Arg Arg Tyr Ser Ser Tyr Arg Ser His Asp  
35 40 45  
His Tyr Gln Arg Gln Arg Val Leu Gln Lys Glu Arg Ala Ile Glu Glu  
50 55 60  
Arg Arg Val Val Phe Ile Gly Lys Ile Pro Gly Arg Met Thr Arg Ser  
65 70 75 80  
Glu Leu Lys Gln Arg Phe Ser Val Phe Gly Glu Ile Glu Glu Cys Thr  
85 90 95  
Ile His Phe Arg Val Gln Gly Asp Asn Tyr Gly Phe Val Thr Tyr Arg  
100 105 110  
Tyr Ala Glu Glu Ala Phe Ala Ala Ile Glu Ser Gly His Lys Leu Arg  
115 120 125  
Gln Ala Asp Glu Gln Pro Phe Asp Leu Cys Phe Gly Gly Arg Arg Xaa  
130 135 140  
Xaa Cys Lys Arg Ser Tyr Ser Asp Leu Asp Ser Asn Arg Arg Glu Asp Phe  
145 150 155 160  
Asp Pro Ala Pro Val Lys Ser Lys Phe Asp Ser Leu Asp Phe Asp Thr  
165 170 175  
Leu Leu Lys Gln Ala Gln Lys Asn Leu Arg Arg  
180 185

(2) INFORMATION FOR SEQ ID NO: 411:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 411:

Met Lys Leu Pro Gly Lys Phe Arg Arg Ala His Gln Gly Asn Leu Glu  
1 5 10 15

Ser Gln Leu Thr Ser Glu Ser Tyr Tyr Lys Glu Thr Leu Ser Val Pro  
 20 25 30  
 5 Thr Val Glu His Ile Ile Gln Glu Leu Lys Asp Ile Phe Ser Glu Gln  
 35 40 45  
 His Leu Lys Ala Leu Lys Cys Leu Ser Leu Val Pro Ser Val Met Gly  
 50 55 60  
 10 Gln Leu Lys Phe Asn Thr Ser Glu His Ala Asp Met Tyr Arg  
 65 70 75 80  
 Ser Asp Leu Pro Asn Pro Asp Thr Leu Ser Ala Glu Leu His Cys Trp  
 85 90 95  
 15 Arg Ile Lys Trp Lys His Arg Gly Lys Asp Ile Glu Leu Pro Ser Thr  
 100 105 110  
 Ile Tyr Glu Ala Leu His Leu Pro Asp Ile Lys Phe Phe Pro Asn Val  
 115 120 125  
 20 Tyr Ala Leu Leu Lys Val Leu Cys Ile Leu Pro Val Met Lys Val Glu  
 130 135 140  
 Asn Glu Arg Tyr Glu Asn Gly Arg Lys Arg Leu Lys Ala Tyr Leu Arg  
 145 150 155 160  
 Asn Thr Leu Thr Asp Gln Arg Ser Ser Asn Leu Ala Leu Leu Asn Ile  
 165 170 175  
 30 Asn Phe Asp Ile Lys His Asp Leu Asp Leu Met Val Asp Thr Tyr Ile  
 180 185 190  
 Lys Leu Tyr Thr Xaa Xaa Ser Xaa Leu Xaa Thr Xaa Xaa Ser Xaa Xaa  
 195 200 205  
 Val Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa  
 210 215 220  
 40 Asp Xaa Xaa Xaa Arg Glu Lys Ala Val Arg Cys Met Xaa  
 225 230 235  
 45 (2) INFORMATION FOR SEQ ID NO: 412:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 412:  
 Met Lys Pro Met Ala Val Val Ala Ser Thr Val Leu Gly Leu Val Gln  
 1 5 10 15  
 55 Asn Met Arg Ala Phe Gly Gly Ile Leu Val Val Val Tyr Tyr Val Phe  
 20 25 30  
 Ala Ile Ile Gly Ile Asn Leu Phe Arg Gly Val Ile Val Ala Leu Pro  
 35 40 45

Gly Asn Ser Ser Leu Ala Pro Ala Asn Gly Ser Ala Pro Cys Gly Ser  
 50 55 60  
 5 Phe Glu Gln Leu Glu Tyr Trp Ala Asn Asn Phe Asp Asp Phe Ala Ala  
 65 70 75 80  
 Ala Leu Val Thr Leu Trp Asn Leu Met Val Val Asn Asn Trp Gln Val  
 85 90 95  
 10 Phe Leu Asp Ala Tyr Arg Arg Tyr Ser Gly Pro Trp Ser Lys Ile Tyr  
 100 105 110  
 Phe Val Leu Trp Trp Leu Val Ser Ser Val Ile Trp Val Asn Leu Phe  
 115 120 125  
 15 Leu Ala Leu Ile Leu Glu Asn Phe Leu His Lys Trp Asp Pro Arg Ser  
 130 135 140  
 20 His Leu Gln Pro Leu Ala Gly Thr Pro Glu Ala Thr Tyr Gln Met Thr  
 145 150 155 160  
 Val Glu Leu Leu Phe Arg Asp Ile Leu Glu Glu Pro Gly Glu Asp Glu  
 165 170 175  
 25 Leu Thr Glu Arg Leu Ser Gln His Pro His Leu Trp Leu Cys Arg Xaa  
 180 185 190  
 30  
 35 (2) INFORMATION FOR SEQ ID NO: 413:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 413:  
 Asn Val Val Val Val Ala Phe Gly Leu Ile Leu Ile Ile Glu Ser Leu  
 1 5 10 15  
 45 Gly Glu Gln Cys Pro  
 20  
 50 (2) INFORMATION FOR SEQ ID NO: 414:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 51 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 414:  
 Met Asn Trp Gly Leu Ser Ile Trp Leu His Tyr Tyr Glu Lys Lys  
 1 5 10 15

Glu Gln Val Phe Leu Val Ile Leu Ala His Val Val Arg Arg Cys Ala  
20 25 30  
Ser Asp Gly Ile Leu Gln Phe Glu Ser Ser Leu Leu Lys Met Arg Arg  
35 40 45  
Ala Pro Xaa  
50

(2) INFORMATION FOR SEQ ID NO: 415:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 415:

Met Leu Ile Ile Ser Leu Arg Pro Gln Phe Pro Ser Leu Ile Val Gln  
1 5 10 15  
Leu Glu Cys Ser Val Leu Phe Leu Pro Ile Ser Leu Asn Leu Leu Leu  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 416:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 163 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 416:

Met Val Lys Val Cys Asn Asp Ser Asp Arg Trp Ser Leu Ile Ser Leu  
1 5 10 15  
Ser Asn Asn Ser Gly Lys Asn Val Glu Leu Lys Phe Val Asp Ser Leu  
20 25 30  
Arg Arg Gln Phe Glu Phe Ser Val Asp Ser Phe Gln Ile Lys Leu Asp  
35 40 45  
Ser Leu Leu Leu Phe Tyr Glu Cys Ser Glu Asn Pro Met Thr Glu Thr  
50 55 60  
Phe His Pro Thr Ile Ile Gly Glu Ser Val Tyr Gly Asp Phe Gln Glu  
65 70 75 80  
Ala Phe Asp His Leu Cys Asn Lys Ile Ile Ala Thr Arg Asn Pro Glu  
85 90 95

Glu Ile Arg Gly Gly Gly Leu Leu Lys Tyr Cys Asn Leu Leu Val Arg  
1 5 10 15 100 110  
Gly Phe Arg Pro Ala Ser Asp Glu Ile Lys Thr Leu Gln Arg Tyr Met

115 120 125  
Cys Ser Arg Phe Phe Ile Asp Phe Ser Asp Ile Gly Glu Gln Gln Arg  
130 135 140  
Lys Leu Glu Ser Tyr Leu Gln Asn His Phe Val Gly Ile Gly Arg Pro  
145 150 155 160  
Gln Val Xaa

(2) INFORMATION FOR SEQ ID NO: 417:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 174 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 417:

Met Ala Pro Lys Gly Lys Val Gly Thr Arg Gly Lys Lys Gln Ile Phe  
1 5 10 15  
Glu Glu Asn Arg Glu Thr Leu Lys Phe Tyr Leu Arg Ile Ile Leu Gly  
20 25 30  
Ala Asn Ala Ile Tyr Cys Leu Val Thr Leu Val Phe Phe Tyr Ser Ser  
35 40 45  
Ala Ser Phe Trp Ala Trp Leu Ala Leu Gly Phe Ser Leu Ala Val Tyr  
50 55 60  
Gly Ala Ser Tyr His Ser Met Ser Ser Met Ala Arg Ala Ala Phe Ser  
65 70 75 80  
Glu Asp Gly Ala Leu Met Asp Gly Gly Met Asp Leu Asn Met Glu Gln  
85 90 95  
Gly Met Ala Glu Glu His Leu Lys Asp Val Ile Leu Leu Thr Ala Ile Val  
100 105 110  
Gln Val Leu Ser Cys Phe Ser Leu Tyr Val Trp Ser Phe Trp Leu Leu  
115 120 125  
Ala Pro Gly Arg Ala Leu Tyr Leu Leu Trp Val Asn Val Leu Gly Pro  
130 135 140  
Trp Phe Thr Ala Asp Ser Gly Thr Pro Ala Pro Glu His Asn Glu Lys  
145 150 155 160  
Arg Gln Arg Arg Gln Glu Arg Arg Gln Met Lys Arg Leu Xaa  
165 170

(2) INFORMATION FOR SEQ ID NO: 418:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 amino acids

60

(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 418:

5 Met Glu Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Val Ala Met Met 15

Arg Pro Phe Tyr Leu Leu Pro Val Leu Cys Thr Gln Ala Leu Arg 30

10 Gln Ser Gln Gly Lys Ser Pro Leu Leu Trp Lys Arg Thr Cys Cys Leu 45

Ala Xaa 50

(2) INFORMATION FOR SEQ ID NO: 419:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 120 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 419:

25 Met Leu Gly Lys Gly Gly Arg Ala Gly Leu Leu Arg Tyr Arg Leu 15

30 Leu Tyr Phe Thr Leu Val Val Gly Glu Gly Glu Pro Gly Glu Asn Lys 30

Val Thr Ile Pro Phe Phe Glu Thr Gly Lys Lys Ile Ile Phe Cys Ser 45

35 Val Lys Met Val Glu Asn Ser Asn Val Pro Ser His Lys Gly Pro Val 60

40 Pro Leu Arg Ser Glu Gln Trp Glu Leu Lys Ile Ser Glu Thr Leu Gly 80

Glu Gly Lys Ile Gly Phe Leu Leu Ile Gly Arg Cys Ser Ser Gly Xaa 95

45 Gly Gly Leu Cys Phe Cys Trp Asp Val Leu Cys Cys Met Tyr Ala Tyr 110

Met Asp Arg Ser Leu Ser Leu 120

(2) INFORMATION FOR SEQ ID NO: 420:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 159 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 420:

60

Met Thr His Leu Leu Leu Thr Ala Thr Val Thr Pro Ser Glu Gln Asn 15

5 Ser Ser Arg Glu Pro Gly Trp Glu Thr Ala Met Ala Lys Asp Ile Leu 30

Gly Glu Ala Gly Leu His Phe Asp Glu Leu Asn Lys Leu Arg Val Leu 45

10 Asp Pro Glu Val Thr Gln Gln Thr Ile Glu Leu Lys Glu Cys Lys 60

Asp Phe Val Asp Lys Ile Gly Gln Phe Gln Lys Ile Val Gly Gly Leu 80

15 Ile Glu Leu Val Asp Gln Leu Ala Lys Glu Ala Glu Asn Glu Lys Met 95

20 Lys Ala Ile Gly Ala Arg Asn Leu Leu Lys Ser Ile Ala Lys Gln Arg 110

Glu Ala Gln Gln Gln Gln Leu Ala Leu Ile Ala Glu Lys Lys Met 125

25 Gln Leu Glu Arg Tyr Arg Val Glu Tyr Glu Ala Leu Cys Lys Val Glu 140

Ala Glu Gln Asn Glu Phe Ile Asp Gln Phe Ile Phe Gln Lys Xaa 155

(2) INFORMATION FOR SEQ ID NO: 421:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 154 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 421:

40 Met Asn Val Gly Val Ala His Ser Glu Val Asn Pro Asn Thr Arg Val 15

45 Met Asn Ser Arg Gly Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu 30

Leu His Ile Val Leu Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala 45

50 Trp Thr Leu Thr Asn Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu 60

His Ala Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala 80

55 Arg Leu Leu Thr His Trp Glu Gln Leu Asp Tyr Gly Val Gln Phe Thr 95

Ser Ser Arg Lys Phe Phe Thr Ile Ser Pro Ile Ile Leu Tyr Phe Leu 110

Ala Ser Phe Tyr Thr Lys Tyr Asp Pro Thr His Phe Ile Leu Asn Thr  
115 120 125

5 Ala Ser Leu Leu Ser Val Leu Ile Pro Lys Met Pro Gln Leu His Gly  
130 135 140

Val Arg Ile Phe Gly Ile Asn Lys Tyr Xaa  
145 150

(2) INFORMATION FOR SEQ ID NO: 422:

15

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 204 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 422:

20

Met Val Cys Gly Gly Phe Ala Cys Ser Lys Asn Cys Leu Cys Ala Leu  
1 5 10 15

25

Asn Leu Leu Tyr Thr Leu Val Ser Leu Leu Leu Ile Gly Ile Ala Ala  
20 25 30

Trp Gly Ile Gly Phe Gly Leu Ile Ser Ser Leu Arg Val Val Gly Val  
35 40 45

30

Val Ile Ala Val Val Lys His Gln Val Leu Leu Phe Phe Tyr Met Ile  
50 55 60

Ile Gly Ala Val Lys His His Gln Val Leu Leu Phe Phe Tyr Met Ile  
65 70 75 80

35

Ile Leu Leu Leu Val Phe Ile Val Gln Phe Ser Val Ser Cys Ala Cys  
85 90 95

40

Leu Ala Leu Asn Gln Gln Gln Gly Gln Leu Leu Glu Val Gly Trp  
100 105 110

Asn Asn Thr Ala Ser Ala Arg Asn Asp Ile Gln Arg Asn Leu Asn Cys  
115 120 125

45

Cys Gly Phe Arg Ser Val Asn Pro Asn Asp Thr Cys Leu Ala Ser Cys  
130 135 140

50

Val Lys Ser Asp His Ser Cys Ser Pro Cys Ala Pro Ile Ile Gly Glu  
145 150 155 160

Tyr Ala Gly Glu Val Leu Arg Phe Val Gly Gly Ile Gly Leu Phe Phe  
165 170 175

55

Ser Phe Thr Glu Ile Leu Gly Val Trp Leu Thr Tyr Arg Tyr Arg Asn  
180 185 190

Gln Lys Asp Pro Arg Ala Asn Pro Ser Ala Phe Leu  
195 200

60

(2) INFORMATION FOR SEQ ID NO: 423:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 67 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 423:

10

Met Leu Gln Ser Ile Ile Lys Asn Ile Trp Ile Pro Met Lys Pro Tyr  
1 5 10 15

Tyr Thr Lys Val Tyr Gln Glu Ile Trp Ile Gly Met Gly Leu Met Gly  
20 25 30

15

Phe Ile Val Tyr Lys Ile Arg Ala Ala Asp Lys Arg Ser Lys Ala Leu  
35 40 45

20

Lys Ala Ser Ala Pro Ala Pro Gly His His Asn Gln Ile Tyr Leu Glu  
50 55 60

Tyr Met Xaa  
65

25

(2) INFORMATION FOR SEQ ID NO: 424:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 424:

35

Met Leu Gly Val Ser Leu Phe Leu Leu Val Val Leu Tyr His Tyr Val  
1 5 10 15

40

Ala Val Asn Asn Pro Lys Lys Gln Glu  
20 25

(2) INFORMATION FOR SEQ ID NO: 425:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 299 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 425:

50

Met Ala Ala Xaa Glu Pro Ala Val Leu Ala Leu Pro Asn Ser Gly Ala  
1 5 10 15

55

Gly Gly Ala Gly Ala Pro Ser Gly Thr Val Pro Val Leu Phe Cys Phe  
20 25 30

Ser Val Phe Ala Arg Pro Ser Ser Val Pro His Gly Ala Gly Tyr Glu  
35 40 45

60

Leu Leu Ile Gln Lys Phe Leu Ser Leu Tyr Gly Asp Gln Ile Asp Met

610

609

50 55 60

His Arg Lys Phe Val Val Gln Leu Phe Ala Glu Glu Thr Gly Gln Tyr 80  
65

Val Asp Leu Pro Lys Gly Phe Ala Val Ser Glu Arg Cys Lys Val Arg 95  
85

Leu Val Pro Leu Gln Ile Gln Leu Thr Thr Leu Gly Asn Leu Thr Pro 110  
100

Ser Ser Thr Val Phe Phe Cys Asp Met Gln Glu Arg Phe Arg Pro 125  
115

Ala Ile Lys Tyr Phe Gly Asp Ile Ile Ser Val Gly Gln Arg Leu Leu 140  
130 135

Gln Gly Ala Arg Ile Leu Gly Ile Pro Val Ile Val Thr Glu Gln Tyr 160  
145 150

Pro Lys Gly Leu Gly Ser Thr Val Gln Glu Ile Asp Leu Thr Gly Val 175  
165 170

Lys Leu Val Leu Pro Lys Thr Lys Phe Ser Met Val Leu Pro Glu Val 190  
180 185

Glu Ala Ala Leu Ala Glu Ile Pro Gly Val Arg Ser Val Val Leu Phe 205  
195

Gly Val Glu Thr His Val Cys Ile Gln Gln Thr Ala Leu Glu Leu Val 220  
210 215

Gly Arg Gly Val Glu Val His Ile Val Ala Asp Ala Thr Ser Ser Arg 240  
225 230 235

Ser Met Met Asp Arg Met Phe Ala Leu Glu Arg Leu Ala Xaa Gly 255  
245 250

Ile Ile Val Thr Thr Ser Glu Ala Val Leu Leu Gln Leu Val Ala Asp 270  
260 265

Lys Asp His Pro Lys Phe Lys Glu Ile Gln Asn Leu Ile Lys Ala Ser 285  
275 280

Ala Pro Glu Ser Gly Leu Leu Ser Lys Val Xaa 295  
290

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 426:

Met Arg Asp Leu Gly Thr Leu Leu Ser Pro Val Cys Ser 10  
1

(2) INFORMATION FOR SEQ ID NO: 427:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 198 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 427:

Met Phe Gly Cys Leu Val Ala Gly Arg Leu Val Gln Thr Ala Ala Gln 15  
1 5 10

Gln Val Ala Glu Asp Lys Phe Val Phe Asp Leu Pro Asp Tyr Glu Ser 30  
20 25

Ile Asn His Val Val Val Phe Met Leu Gly Thr Ile Pro Phe Pro Glu 45  
35 40

Gly Met Gly Gly Ser Val Tyr Phe Ser Tyr Pro Asp Ser Asn Gly Met 60  
50 55

Pro Val Trp Gln Leu Leu Gly Phe Val Thr Asn Gly Lys Pro Ser Ala 80  
65 70 75

Ile Phe Lys Ile Ser Gly Leu Lys Ser Gly Glu Gly Ser Gln His Pro 95  
85 90

Phe Gly Ala Met Asn Ile Val Arg Thr Pro Ser Val Ala Gln Ile Gly 110  
100 105

Ile Ser Val Glu Leu Leu Asp Ser Met Ala Gln Gln Thr Pro Val Gly 125  
115 120

Asn Ala Ala Val Ser Ser Val Asp Ser Phe Thr Gln Phe Thr Gln Lys 140  
130 135

Met Leu Asp Asn Phe Tyr Asn Phe Ala Ser Ser Phe Ala Val Ser Gln 160  
145 150 155

Ala Gln Met Thr Pro Ser Pro Ser Glu Met Phe Ile Pro Ala Asn Val 175  
165 170

Val Leu Lys Trp Tyr Glu Asn Phe Gln Arg Arg Leu Ala Gln Asn Pro 190  
180 185

Xaa Phe Trp Xaa Thr Xaa 195  
190

(2) INFORMATION FOR SEQ ID NO: 428:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 428:

Met Gly Leu Pro Leu Met Ala Leu Met Trp Ser Thr Leu Pro Ala Ser 10

1 5 10 15  
Ala Gly Val Asn Phe Ile Leu Ala Leu Pro Leu Leu Leu Trp Lys  
20 25 30  
5 Asn Arg Gly Val Gly Arg Ser Val Met Ser Ala Val Gly Xaa  
35 40 45  
10 (2) INFORMATION FOR SEQ ID NO: 429:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 370 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 429:  
Met Lys Lys Val Gly Glu Lys Arg Val Asp Val Asn Ser Ala Val Ala  
1 5 10 15  
Met Gly Glu Val Ile Leu Ala Val Cys His Pro Asp Cys Ile Thr Thr  
20 25 30  
25 Ile Lys His Trp Ile Thr Ile Ile Arg Ala Arg Phe Gly Glu Val Leu  
35 40 45  
Thr Trp Ala Lys Gln His Gln Arg Leu Glu Thr Ala Leu Ser Glu  
50 55 60  
Leu Val Ala Asn Ala Glu Leu Leu Glu Leu Ala Trp Ile Gln  
65 70 75 80  
30 Trp Ala Glu Thr Thr Leu Ile Gln Arg Asp Gln Glu Pro Ile Pro Gln  
85 90 95  
35 Asn Ile Asp Arg Val Lys Ala Leu Ile Ala Glu His Gln Thr Phe Met  
100 105 110  
Glu Glu Met Thr Arg Lys Gln Pro Asp Val Asp Arg Val Thr Lys Thr  
115 120 125  
40 Tyr Lys Arg Lys Asn Ile Glu Pro Thr His Ala Pro Phe Ile Glu Lys  
130 135 140  
45 Ser Arg Ser Gly Arg Lys Ser Leu Ser Gln Pro Thr Pro Pro Pro  
145 150 155 160  
Met Pro Ile Leu Ser Gln Ser Glu Ala Lys Asn Pro Arg Ile Asn Gln  
165 170 175  
50 Leu Ser Ala Arg Trp Gln Gln Val Trp Leu Leu Ala Leu Glu Arg Gln  
180 185 190  
55 Arg Lys Leu Asn Asp Ala Leu Asp Arg Leu Glu Leu Lys Glu Phe  
195 200 205  
Ala Asn Phe Asp Phe Asp Val Trp Arg Lys Lys Tyr Met Arg Trp Met  
210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995

Asn His Lys Lys Ser Arg Val Met Asp Phe Phe Arg Arg Ile Asp Lys  
225 230 235 240  
5 Asp Gln Asp Gly Lys Ile Thr Arg Gln Glu Phe Ile Asp Gly Ile Leu  
245 250 255  
Ala Ser Lys Phe Pro Thr Thr Lys Leu Glu Met Thr Ala Val Ala Asp  
260 265 270  
10 Ile Phe Asp Arg Asp Gly Asp Gly Tyr Ile Asp Tyr Tyr Glu Phe Val  
275 280 285  
Ala Ala Leu His Pro Asn Lys Asp Ala Tyr Arg Pro Thr Thr Asp Ala  
290 295 300  
15 Asp Lys Ile Glu Asp Glu Val Thr Arg Gln Val Ala Gln Cys Lys Cys  
305 310 315 320  
Ala Lys Arg Phe Gln Val Glu Gln Ile Gly Glu Asn Lys Tyr Arg Phe  
325 330 335  
Phe Leu Gly Asn Gln Phe Gly Asp Ser Gln Gln Leu Leu Val Arg  
340 345 350  
25 Ile Leu Arg Asn Arg Asp Gly Ser Arg Trp Trp Arg Met Asp Gly Leu  
355 360 365  
Gly Xaa  
370  
30 (2) INFORMATION FOR SEQ ID NO: 430:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 430:  
Met Asn Val Lys Thr Phe Ser Xaa Asp His Met His Phe Leu Cys Cys  
1 5 10 15  
45 Leu Tyr Leu Arg Tyr Val Thr Phe Val Tyr Leu Asn Leu Phe  
20 25 30  
50 (2) INFORMATION FOR SEQ ID NO: 431:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 431:  
Met Glu Pro His Leu Arg Cys Arg Val Thr Arg Val Arg Gly Ser Leu  
1 5 10 15  
60 Gly Asn Thr Gly Arg Trp Leu Leu

613

614

20

5 (2) INFORMATION FOR SEQ ID NO: 412:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 412:

Met His Tyr Leu Val Leu Gly Gly Leu Gly Val Phe Leu Phe Ser  
1 5 10 1515 Cys Phe Val Phe Leu Phe Phe Xaa Phe Ser Phe Ala Phe Phe Pro Phe  
20 25 3020 Tyr Leu Glu Gly Met Gly Gly Ser Gly Asn Arg Glu Val Gly Gly Gly  
35 40 45Phe Cys Leu Phe Phe  
50

25

(2) INFORMATION FOR SEQ ID NO: 413:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 176 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 413:

35 Met Val Ser Lys Ala Leu Leu Arg Leu Val Ser Ala Val Asn Arg Arg  
1 5 10 1540 Arg Met Lys Leu Leu Leu Gly Ile Ala Leu Leu Ala Tyr Val Ala Ser  
20 25 30Val Trp Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu Asn Gly Glu  
35 40 4545 Leu Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu Pro Leu Arg Glu  
50 55 60Lys Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val  
65 70 75 8050 Lys Phe Leu Ser Glu Lys Asp Arg Lys Arg Ile Leu Ile Thr Gly Gly  
85 90 95Ala Gly Phe Val Gly Ser His Leu Thr Asp Lys Leu Met Met Asp Gly  
100 105 11055 His Glu Val Thr Val Val Asp Asn Phe Phe Thr Gly Arg Lys Arg Asn  
115 120 12560 Val Glu His Trp Ile Gly His Glu Asn Phe Glu Leu Ile Asn His Asp  
130 135 140Val Trp Ser Pro Ser Thr Arg Leu Thr Arg Tyr Thr Ile Trp His  
145 150 155 1605 Leu Gln Pro Pro Leu Gln Thr Thr Cys Ile Ile Leu Ser Arg His Xaa  
165 170 175

10

(2) INFORMATION FOR SEQ ID NO: 434:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 434:

20 Met Leu Arg Cys Trp Pro Leu Phe Trp Leu Pro Leu Val Ser Pro Phe  
1 5 10 1525 Cys Ser Leu Phe Trp Leu Leu Val Glu Trp Phe Gly Thr Asn Ile Asp  
20 25 30Arg Glu Ser Tyr Asp Ala Ile Gly Gly Pro Ser Trp Met Thr Ala Ser  
35 40 4530 Ser Phe Cys Leu Ser Asn Ser Asn Ile Trp Ser Leu Glu Ile Ser Ser  
50 55 6035 Gly Ser Thr Ser Val Val His Ser Gln Gln Ala Met Asp  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 435:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 435:

45 Met Arg Ser Cys Glu Ile Gln Leu Cys Val Trp Leu Leu Val Ser Ser  
1 5 10 1550 His Val Asp Met Val Leu Gly Gly Ser Pro Ser Thr Leu Tyr Met Met  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 436:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids



(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 436:

5 Met Val Val Asn Ser Leu Cys Phe Leu Ser Leu Leu Val Ile Leu  
1 5 10 15  
Glu Leu Ser Thr Asp Ser Ser Ala Arg Leu Leu Tyr His Glu  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 437:

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 69 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 437:

20 Met Asp Lys Glu Lys His Leu Glu Val Arg Arg Ser Val Phe Lys Ile  
1 5 10 15  
Gln Gly Lys Ile Ala Phe Ser Leu Met Phe Val Leu Lys Asp Leu Ser  
20 25 30  
Pro Thr Ile Phe Ser His Ser Ile Leu Leu Leu Leu Pro His His Val  
35 40 45  
Leu Pro Cys Thr Pro Gln Met Val Arg Gly Val Thr Gln Val Leu Arg  
50 55  
Glu Phe Gly Asp Gln  
65

(2) INFORMATION FOR SEQ ID NO: 438:

40 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 438:

45 Met Pro Leu Cys Phe Phe Ser Phe Leu Cys Trp Val Leu Val Phe  
1 5 10 15  
Lys Leu Ile  
50

(2) INFORMATION FOR SEQ ID NO: 439:

55 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 439:

Met Lys Phe Ser Leu Val Leu Leu Ile Lys Ile Ile Ser Phe Glu Arg  
1 5 10 15

5 Leu Leu Ile Phe Leu Phe Pro Leu Ser Phe Leu Pro Asn Ile Trp Arg  
20 25 30

Arg Val Met Val Asn Leu Asn Ile Leu Phe Xaa  
35 40

(2) INFORMATION FOR SEQ ID NO: 440:

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 440:

20 Met Leu Leu Phe Pro Ser Leu Leu Phe Ala Ala Thr Tyr Asn Val Ala  
1 5 10 15  
Asn Pro Ser Arg Leu Ile Leu Tyr Met Ile Ser Ala Gly Ala Asp Ser  
20 25 30  
Gln

(2) INFORMATION FOR SEQ ID NO: 441:

30 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 53 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 441:

40 Met Trp Gln Val Arg Gly Leu Pro Pro Val Pro Leu Leu Thr Met  
1 5 10 15  
Ser Pro Pro Cys Leu Ser Ser Pro Phe Pro Phe Ile Ser Val Pro  
20 25 30  
Leu Phe Glu Ala Val Pro Ile Ser Val Ser Asp Gln Pro Ser Pro Xaa  
35 40 45

Leu Thr Thr Leu Leu  
50

(2) INFORMATION FOR SEQ ID NO: 442:

55 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 442:

60

Met Ile Thr Ser Val Leu Val Phe Leu Ile Phe Phe Pro Tyr Leu  
1 5 10 15  
5 Ser Leu Val Thr Leu Leu Gln Ala Arg Asn Leu Trp Val Ile His Arg  
20 25 30  
Ala Ala Leu Cys Glu Ser Gly Leu Phe His Trp Arg Lys Gly Ile Glu  
35 40 45  
10 Asn Gln Leu Glu Pro Met Tyr Phe Leu Pro His Gly Thr Leu Phe Leu  
50 55 60

15

## (2) INFORMATION FOR SEQ ID NO: 443:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 443:

Met Leu Tyr Ser Cys Glu Pro Tyr Leu Ile Ile Leu Asn Ile Tyr Ser  
1 5 10 15

30 Gln Lys Ala Phe Tyr Phe Tyr Phe Glu Gly Ser Phe Ser Val Cys  
20 25 30

Thr Leu

35

## (2) INFORMATION FOR SEQ ID NO: 444:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 89 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 444:

Met Arg Gln Arg Gln Ala Ala Cys Gln Pro Pro Ser Arg Asn Gly  
1 5 10 15

50 Leu Ala Gln Glu Cys Pro Pro His Ile Pro Ser Ser Phe Leu Val  
20 25 30

Lys Leu Leu Phe Ile Pro Trp Leu Ala Ser Leu Leu Ser Ser Pro Leu  
35 40 45

55 Asn Leu Leu Leu Val Ser Ile Ser Trp Asp Leu Gly Leu Lys Leu  
50 55 60

Asn Leu Gln Cys Arg Gln His Gln Val Leu Gln Glu Lys Asn Thr  
65 70 75 80

60

Lys Lys Phe Asn Lys Lys Lys Lys Lys  
85

5

## (2) INFORMATION FOR SEQ ID NO: 445:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 350 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 445:

Met Asp Phe Ile Thr Ser Thr Ala Ile Leu Pro Leu Leu Phe Gly Cys  
1 5 10 15

Leu Gly Val Phe Gly Leu Phe Arg Leu Leu Gln Trp Val Arg Gly Lys  
20 25 30

20 Ala Tyr Leu Arg Asn Ala Val Val Ile Thr Gly Ala Thr Ser Gly  
35 40 45

Leu Gly Lys Glu Cys Ala Lys Val Phe Tyr Ala Ala Gly Ala Lys Leu  
50 55 60

25 Val Leu Cys Gly Arg Asn Gly Gly Ala Leu Glu Glu Leu Ile Arg Glu  
65 70 75 80

Leu Thr Ala Ser His Ala Thr Lys Val Gln Thr His Lys Pro Tyr Leu  
85 90 95

Val Thr Phe Asp Leu Thr Asp Ser Gly Ala Ile Val Ala Ala Ala  
100 105 110

35 Glu Ile Leu Gln Cys Phe Gly Tyr Val Asp Ile Leu Val Asn Asn Ala  
115 120 125

Gly Ile Ser Tyr Arg Gly Thr Ile Met Asp Thr Thr Val Asp Val Asp  
130 135 140

Lys Arg Val Met Glu Thr Asn Tyr Phe Gly Pro Val Ala Leu Thr Lys  
145 150 155 160

Ala Leu Leu Pro Ser Met Ile Lys Arg Arg Gln Gly His Ile Val Ala  
165 170 175

Ile Ser Ser Ile Gln Gly Lys Met Ser Ile Pro Phe Arg Ser Ala Tyr  
180 185 190

50 Ala Ala Ser Lys His Ala Thr Gln Ala Phe Phe Asp Cys Leu Arg Ala  
195 200 205

Glu Met Glu Gln Tyr Glu Ile Glu Val Thr Val Ile Ser Pro Gly Tyr  
210 215 220

55 Ile His Thr Asn Leu Ser Val Asn Ala Ile Thr Ala Asp Gly Ser Arg  
225 230 235 240

Tyr Gly Val Met Asp Thr Thr Thr Ala Gln Gly Arg Ser Pro Val Glu  
245 250 255

60

Val Ala Gln Asp Val Leu Ala Val Gly Lys Lys Lys Asp Val  
260 265 270  
Ile Leu Ala Asp Leu Leu Pro Ser Leu Ala Val Tyr Leu Arg Thr Leu  
275 280 285  
Ala Pro Gly Leu Phe Ser Leu Met Pro Pro Gly Pro Glu Lys Ser  
290 295 300  
Gly Asn Pro Arg Thr Pro Ser Thr Leu Thr Ser Gln Gly Gln Gly Arg  
305 310 315 320  
Glu Ala Ala Leu Leu Gly Leu Leu Thr Leu Gln Gly Thr Val Ala Phe  
325 330 335  
Val Glu Thr Leu Met Glu Ile Cys Leu Thr Ser Gly Lys Asp  
340 345 350

(2) INFORMATION FOR SEQ ID NO: 446:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 446:

Met Val Phe Leu Pro Arg Gly Val Val Ser Gly Gly Ala Ala Cys  
1 5 10 15  
Leu Trp Leu Thr Phe Ile Leu Glu Thr Glu Val Tyr Leu Asp Leu Ala  
20 25 30  
Thr Glu Ala Arg Ala His Ser Arg Met Gly Leu Gly Leu Trp Pro Pro  
35 40 45

40 Asn

(2) INFORMATION FOR SEQ ID NO: 447:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 278 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 447:

Met Ala Ser Ala Glu Leu Asp Tyr Thr Ile Glu Ile Pro Asp Gln Pro  
1 5 10 15  
Cys Trp Ser Gln Lys Asn Ser Pro Ser Pro Gly Gly Lys Glu Ala Glu  
20 25 30  
Thr Arg Gln Pro Val Val Ile Leu Leu Gly Trp Gly Gly Cys Lys Asp  
35 40 45 50 55 60

Lys Asn Leu Ala Lys Tyr Ser Ala Ile Tyr His Lys Arg Gly Cys Ile  
50 55 60  
Val Ile Arg Tyr Thr Ala Pro Trp His Met Val Phe Phe Ser Glu Ser  
65 70 75 80  
Leu Gly Ile Pro Ser Leu Arg Val Leu Ala Gln Lys Leu Leu Glu Leu  
85 90 95  
Leu Phe Asp Tyr Glu Ile Glu Lys Glu Pro Leu Leu Phe His Val Phe  
100 105 110  
Ser Asn Gly Gly Val Met Leu Tyr Arg Tyr Val Leu Glu Leu Gln  
115 120 125  
Thr Arg Arg Phe Cys Arg Leu Arg Val Val Gly Thr Ile Phe Asp Ser  
130 135 140  
Ala Pro Gly Asp Ser Asn Leu Val Gly Ala Leu Arg Ala Ala Ala  
145 150 155 160  
Ile Leu Glu Arg Arg Ala Ala Met Leu Arg Leu Leu Leu Val Ala  
165 170 175  
Phe Ala Leu Val Val Val Leu Phe His Val Leu Leu Ala Pro Ile Thr  
180 185 190  
Ala Xaa Phe His Thr His Phe Tyr Asp Arg Leu Gln Asp Ala Gly Ser  
195 200 205  
Arg Trp Pro Glu Leu Tyr Leu Tyr Ser Arg Ala Asp Glu Val Val Leu  
210 215 220  
Ala Arg Asp Ile Glu Arg Met Val Glu Ala Arg Leu Ala Arg Arg Val  
225 230 235 240  
Leu Ala Arg Ser Val Asp Phe Val Ser Ser Ala His Val Ser His Leu  
245 250 255  
Arg Asp Tyr Pro Thr Tyr Tyr Thr Ser Leu Cys Val Asp Phe Met Arg  
260 265 270

Asn Cys Val Arg Cys Xaa  
275

(2) INFORMATION FOR SEQ ID NO: 448:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 199 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 448:

Met Ser Phe Ile Phe Asp Trp Ile Tyr Ser Gly Phe Ser Ser Val Leu  
1 5 10 15  
Gln Phe Leu Gly Leu Tyr Lys Lys Thr Gly Lys Leu Val Phe Leu Gly  
20 25 30 35 40 45 50 55 60

621

Leu Asp Asn Ala Gly Lys Thr Thr Leu Leu His Met Leu Lys Asp Asp  
35 40 45

5 Arg Leu Gly Gln His Val Pro Thr Leu His Pro Thr Ser Glu Glu Leu  
50 55 60

Thr Ile Ala Gly Met Thr Thr Thr Phe Asp Leu Gly Gly His Val  
65 70 75 80

10 Gln Ala Arg Arg Val Trp Lys Asn Tyr Leu Pro Ala Ile Asn Gly Ile  
85 90 95

15 Val Phe Leu Val Asp Cys Ala Asp His Glu Arg Leu Leu Glu Ser Lys  
100 105 110

Glu Glu Leu Asp Ser Leu Met Thr Asp Glu Thr Ile Ala Asn Val Pro  
115 120 125

20 Ile Leu Ile Leu Gly Asn Lys Ile Asp Arg Pro Glu Ala Ile Ser Glu  
130 135 140

Glu Arg Leu Arg Glu Met Phe Gly Leu Tyr Gly Gln Thr Thr Gly Lys  
145 150 155 160

25 Gly Ser Ile Ser Leu Lys Glu Leu Asn Ala Arg Pro Leu Glu Val Phe  
165 170 175

30 Met Cys Ser Val Leu Lys Arg Gln Gly Tyr Gly Glu Gly Phe Arg Trp  
180 185 190

Met Ala Gln Tyr Ile Asp Xaa  
195

35

(2) INFORMATION FOR SEQ ID NO: 449:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 258 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 449:

45 Met Thr Leu Ser Arg Phe Ala Tyr Asn Gly Lys Arg Cys Pro Ser Ser  
1 5 10 15

Tyr Asn Ile Leu Asp Asn Ser Lys Ile Ile Ser Glu Glu Cys Arg Lys  
20 25 30

50 Glu Leu Thr Ala Leu Leu His His Tyr Tyr Pro Ile Glu Ile Asp Pro  
35 40 45

55 His Arg Thr Val Lys Glu Lys Leu Pro His Met Val Glu Trp Trp Thr  
50 55 60

Lys Ala His Asn Leu Leu Cys Gln Gln Lys Ile Gln Lys Phe Gln Ile  
65 70 75 80

60 Ala Gln Val Val Arg Glu Ser Asn Ala Met Leu Arg Glu Gly Tyr Lys

WO 98/39448

622

85 90 95

Thr Phe Phe Asn Thr Leu Tyr His Asn Asn Ile Pro Leu Phe Ile Phe  
100 105 110

5 Ser Ala Gly Ile Gly Asp Ile Leu Glu Glu Ile Ile Arg Gln Met Lys  
115 120 125

10 Val Phe His Pro Asn Ile His Ile Val Ser Asn Tyr Met Asp Phe Asn  
130 135 140

Glu Asp Gly Phe Leu Gln Gly Phe Lys Gly Gln Leu Ile His Thr Tyr  
145 150 155 160

15 Asn Lys Asn Ser Ser Val Cys Glu Asn Xaa Gly Tyr Phe Gln Gln Leu  
165 170 175

Glu Gly Lys Thr Asn Val Ile Leu Leu Gly Asp Ser Ile Gly Asp Leu  
180 185 190

20 Thr Met Ala Asp Gly Val Pro Gly Val Gln Asn Ile Leu Lys Ile Gly  
195 200 205

Phe Leu Asn Asp Lys Val Glu Glu Arg Arg Xaa Arg Tyr Met Asp Ser  
210 215 220

Tyr Asp Ile Val Leu Glu Lys Asp Glu Thr Leu Asp Val Val Asn Gly  
225 230 235 240

30 Leu Leu Gln His Ile Leu Cys Gln Gly Val Gln Leu Glu Met Gln Gly  
245 250 255

Pro Xaa

35

(2) INFORMATION FOR SEQ ID NO: 450:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 87 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 450:

45 Met Ser His Val Leu Leu Cys Pro Ser Leu Ser Cys Ser Asn Leu Leu  
1 5 10 15

Pro Pro Ser His Ser Leu Gly Thr Met Gly Ser Leu Ser Pro His Leu  
20 25 30

Cys Gly His Thr Met Cys Pro Val Asn Pro Glu Leu Pro Leu Ser Ser  
35 40 45

55 Arg Leu Thr Thr Asp Gln Pro Gln Pro Asp Ala Cys Ser Pro Thr Leu  
50 55 60

Leu Thr Leu Pro Leu Pro Ser Ser Phe Leu Pro His Ser Lys Pro Thr  
65 70 75 80

Phe Xaa His Pro Cys Ser Pro  
85

5 (2) INFORMATION FOR SEQ ID NO: 451:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 315 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 451:

15 Met Phe Ser Ile Asn Pro Leu Glu Asn Leu Lys Val Tyr Ile Ser Ser  
1 5 10 15

Arg Pro Pro Leu Val Val Phe Met Ile Ser Val Xaa Pro Met Ala Ile  
20 25 30

20 Ala Phe Leu Thr Leu Gly Tyr Phe Phe Lys Ile Lys Glu Ile Lys Ser  
35 40 45

Pro Glu Met Ala Glu Asp Trp Asn Thr Phe Leu Leu Arg Phe Asn Asp  
50 55 60

25 Leu Asp Leu Cys Val Ser Glu Asn Glu Thr Leu Lys His Leu Thr Asn  
65 70 75 80

30 Asp Thr Thr Thr Pro Glu Ser Thr Met Thr Ser Gly Gln Ala Arg Ala  
85 90 95

Ser Thr Gln Ser Pro Gln Ala Leu Glu Asp Ser Gly Pro Val Asn Ile  
100 105 110

35 Ser Val Ser Ile Thr Leu Thr Leu Asp Pro Leu Lys Pro Phe Gly Gly  
115 120 125

40 Tyr Ser Arg Asn Val Thr His Leu Tyr Ser Thr Ile Leu Gly His Gln  
130 135 140

Ile Gly Leu Ser Gly Arg Glu Ala His Glu Glu Ile Asn Ile Thr Phe  
145 150 155 160

45 Thr Leu Pro Thr Ala Trp Ser Ser Asp Asp Cys Ala Leu His Gly His  
165 170 175

Cys Glu Gln Val Val Phe Thr Ala Cys Met Thr Leu Thr Ala Ser Pro  
180 185 190

50 Gly Val Phe Pro Val Thr Val Gln Pro Pro His Cys Val Pro Asp Thr  
195 200 205

Tyr Ser Asn Ala Thr Leu Trp Tyr Lys Ile Phe Thr Thr Ala Arg Asp  
210 215 220

55 Ala Asn Thr Lys Tyr Ala Gln Asp Tyr Asn Pro Phe Trp Cys Tyr Lys  
225 230 235 240

60 Gly Ala Ile Gly Lys Val Tyr His Ala Leu Asn Pro Lys Leu Thr Val  
245 250 255

Ile Val Pro Asp Asp Asp Arg Ser Leu Ile Asn Leu His Leu Met His  
260 265 270

5 Thr Ser Tyr Phe Leu Phe Val Met Val Ile Thr Met Phe Cys Tyr Ala  
275 280 285

10 Val Ile Lys Gly Arg Pro Ser Lys Leu Arg Gln Ser Asn Pro Glu Phe  
290 295 300

Cys Pro Glu Lys Val Ala Leu Ala Glu Ala Xaa  
305 310 315

15 (2) INFORMATION FOR SEQ ID NO: 452:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 452:

25 Met Pro Gly Leu Ser Leu Ala Leu Leu Pro Phe Gly Pro Gly Cys Thr  
1 5 10 15

Glu Ala Leu His Ala Gly Cys Phe Pro Ala Phe Ala Ser Ala Thr Arg  
20 25 30

30 Val Asn Gly Glu Ala Ala Leu Ser Pro Gly Leu Cys Asp Pro Ile Ser  
35 40 45

Val Pro Tyr Val  
50

35 (2) INFORMATION FOR SEQ ID NO: 453:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 383 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 453:

45 Met Ala Val Gly Gln Ile Met Thr Phe Gly Ser Pro Val Ile Gly Cys  
1 5 10 15

50 Gly Phe Ile Ser Gly Trp Asn Leu Val Ser Met Cys Val Glu Tyr Val  
20 25 30

Leu Leu Trp Lys Val Tyr Gln Lys Thr Pro Ala Leu Ala Val Lys Ala  
35 40 45

55 Gly Leu Lys Glu Glu Glu Thr Glu Leu Lys Gln Leu Asn Leu His Lys  
50 55 60

60 Asp Thr Glu Pro Lys Pro Leu Glu Gly Thr His Leu Met Gly Val Lys  
65 70 75 80

625

Asp Ser Asn Ile His Glu Leu Glu His Glu Gln Glu Pro Thr Cys Ala 95  
 85  
 5 Ser Gln Met Ala Glu Pro Phe Arg Thr Phe Arg Asp Gly Trp Val Ser 110  
 100  
 Tyr Tyr Asn Gln Pro Val Phe Leu Ala Gly Met Gly Leu Ala Phe Leu 125  
 115  
 10 Tyr Met Thr Val Leu Gly Phe Asp Cys Ile Thr Thr Gly Tyr Ala Tyr 140  
 130  
 Thr Gln Gly Leu Ser Gly Phe His Pro Gln Tyr Phe Asp Gly Ser Ile 160  
 145  
 15 Ser Tyr Asn Trp Asn Asn Gly Asn Cys Ser Phe Tyr Leu Ala Thr Ser 175  
 165  
 Lys Met Trp Phe Gly Ser Ala Gly Leu Ile Ser Gly Leu Ala Gln Leu 190  
 180  
 20 Ser Cys Leu Ile Leu Cys Val Ile Ser Val Phe Met Pro Gly Ser Pro 205  
 195  
 25 Leu Asp Leu Ser Val Ser Pro Phe Glu Asp Ile Arg Ser Arg Phe Ile 220  
 210  
 Gln Gly Glu Ser Ile Thr Pro Thr Lys Ile Pro Glu Ile Thr Thr Glu 240  
 225  
 30 Ile Tyr Met Ser Asn Gly Ser Asn Ser Ala Asn Ile Val Pro Glu Thr 255  
 245  
 Ser Pro Glu Ser Val Pro Ile Ile Ser Val Ser Leu Leu Phe Ala Gly 270  
 260  
 Val Ile Ala Ala Arg Ile Gly Leu Trp Ser Phe Asp Leu Thr Val Thr 285  
 275  
 40 Gln Leu Leu Gln Glu Asn Val Ile Glu Ser Glu Arg Gly Ile Ile Asn 300  
 290  
 Gly Val Gln Asn Ser Met Asn Tyr Leu Leu Asp Leu Leu His Phe Ile 320  
 305  
 45 Met Val Ile Leu Ala Pro Asn Pro Glu Ala Phe Gly Leu Leu Val Leu 335  
 325  
 Ile Ser Val Ser Phe Val Ala Met Gly His Ile Met Tyr Phe Arg Phe 350  
 340  
 Ala Gln Asn Thr Leu Gly Asn Lys Leu Phe Ala Cys Gly Pro Asp Ala 365  
 355  
 55 Lys Glu Val Arg Lys Glu Asn Gln Ala Asn Thr Ser Val Val Xaa 380  
 370

60 (2) INFORMATION FOR SEQ ID NO: 454:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 186 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 454:

Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu Leu Gly Leu Gly Phe 15  
 1  
 5  
 10 Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe Gly Ser Glu Pro Trp Met 30  
 20  
 15 Met Trp Ala Ala Gly Ala Val Ala Ala Met Ser Ser Ile Thr Phe Pro 45  
 35  
 Ala Val Ser Ala Leu Val Ser Arg Thr Ala Asp Ala Asp Gln Gln Gly 60  
 50  
 20 Val Val Gln Gly Met Ile Thr Gly Ile Arg Gly Leu Cys Asn Gly Leu 80  
 65  
 Gly Pro Ala Leu Tyr Gly Phe Ile Phe Tyr Ile Phe His Val Glu Leu 95  
 85  
 25 Lys Glu Leu Pro Ile Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro 110  
 100  
 Gln His His Phe Glu Gln Asn Ser Ile Ile Pro Gly Pro Pro Phe Leu 125  
 115  
 Phe Gly Ala Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile 140  
 130  
 35 Pro Glu His Thr Asn Leu Ser Leu Arg Ser Ser Trp Arg Lys His 160  
 145  
 Cys Gly Ser His Ser His Pro His Asn Thr Gln Ala Pro Gly Glu Ala 175  
 165  
 40 Lys Glu Pro Leu Leu Gln Asp Thr Asn Val 185  
 180  
 45 (2) INFORMATION FOR SEQ ID NO: 455:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 163 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 455:

Met Leu Gln Thr Ser Asn Tyr Ser Leu Val Leu Ser Leu Gln Phe Leu 15  
 1  
 5  
 10 Leu Leu Ser Tyr Asp Leu Phe Val Asn Ser Phe Ser Glu Leu Leu Gln 30  
 20  
 60 Lys Thr Pro Val Ile Gln Leu Val Leu Phe Ile Ile Gln Asp Ile Ala

626

35 40 45  
Val Leu Phe Asn Ile Ile Ile Phe Leu Met Phe Phe Asn Thr Phe  
50 55 60  
5 Val Phe Gln Ala Gly Leu Val Asn Leu Leu Phe His Lys Phe Lys Gly  
65 70 75 80  
10 Thr Ile Ile Leu Thr Ala Val Tyr Phe Ala Leu Ser Ile Ser Leu His  
85 90 95  
Val Trp Val Met Asn Leu Arg Trp Lys Asn Ser Asn Ser Phe Ile Trp  
100 105 110  
15 Thr Asp Gly Leu Gln Met Leu Phe Val Phe Gln Arg Leu Ala Val  
115 120 125  
Leu Tyr Cys Tyr Phe Tyr Lys Arg Thr Ala Val Arg Leu Gly Asp Pro  
130 135 140  
20 His Phe Tyr Gln Asp Ser Leu Trp Leu Arg Lys Gln Phe Met Gln Val  
145 150 155 160  
Arg Arg Xaa  
25  
(2) INFORMATION FOR SEQ ID NO: 456:  
30 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 46 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 456:  
35 Met Arg Ile Gln Val Phe Ile Leu Leu Leu Gly Ala Gly Gly Thr Ser  
1 5 10 15  
40 Gln Phe Thr Lys Pro Pro Ser Leu Pro Leu Gln Pro Gln Pro Ala Val  
20 25 30  
45 Gln Ser Ser Pro Thr Gln Thr Ser Gln Gln Ile Arg Gln Lys  
35 40 45  
(2) INFORMATION FOR SEQ ID NO: 457:  
50 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 105 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 457:  
55 Met Ser Tyr Leu Ala Phe Leu Tyr Met Thr Phe Asp Phe Cys Cys Leu  
1 5 10 15  
60 Tyr Phe Ser Thr Val Tyr Ala Pro Ser Phe Lys Tyr Ile Cys Val His  
20 25 30

Thr Asp Thr His Ile Cys Val Cys Val Cys Ile Tyr Leu Ser Ser Val  
35 40 45  
5 Val Ser Lys Ser Ser Ala Gln Ala Asp Gly Val Leu Gln Pro Arg Arg  
50 55 60  
10 His Pro Ala Ser Leu Leu Ile Val Phe Ala Thr Ser Ile Ser Gln Ser  
65 70 75 80  
Ser Leu Leu Ile Phe Ser Phe Gln Lys Thr Gln Ala Lys Leu Ile Val  
85 90 95  
15 Phe Ala Val Ser Leu Ala Ala Lys Xaa  
100 105  
(2) INFORMATION FOR SEQ ID NO: 458:  
20 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 70 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 458:  
25 Met Leu Pro Pro Phe Ser Leu Val Tyr Thr His Phe Leu Val Ala Ser  
1 5 10 15  
30 Leu Leu Pro Val Ile Leu Ala Val Phe Pro Asp Ser Ala Gln Ile Val  
20 25 30  
35 Pro Leu Leu Lys Pro Ile Pro Arg Pro Gln Pro Gln Val Ile Phe Pro  
35 40 45  
Ser Ser Gln Leu Leu Gln Gln Leu Leu Ser Val Gln Phe Val Trp Gln  
50 55 60  
40 Ala His Thr Val Ala Xaa  
65 70  
(2) INFORMATION FOR SEQ ID NO: 459:  
45 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 155 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 459:  
50 Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe  
1 5 10 15  
55 Ala Leu Val Gly Leu Ala Lys Leu Ser Gln Gln Ile Ser Ala Pro Val  
20 25 30  
60 Ser Gln Arg Met Asn Ala Leu Phe Val Gln Phe Ala Gln Val Phe Pro  
35 40 45

Leu Lys Val Phe Gly Tyr Gln Pro Asp Pro Leu Asn Tyr Gln Ile Ala  
50 55 60

5 Val Gly Phe Leu Glu Leu Leu Ala Gly Leu Leu Val Met Gly Pro  
65 70 75 80  
Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu Leu Met Met  
85 90 95

10 Gly Ala Ile Phe Thr Leu Ala Leu Lys Glu Ser Leu Ser Thr Cys  
100 105 110

Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Asn Val Gly  
115 120 125

15 Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys  
130 135 140

20 Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys Xaa  
145 150 155

25 (2) INFORMATION FOR SEQ ID NO: 460:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 312 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 460:

Met Lys Leu Gly Arg Ala Val Leu Gly Leu Leu Leu Ala Pro Ser  
1 5 10 15

35 Val Val Gln Ala Val Glu Pro Ile Ser Leu Gly Leu Ala Leu Ala Gly  
20 25 30

40 Val Leu Thr Gly Tyr Ile Tyr Pro Arg Leu Tyr Cys Leu Phe Ala Glu  
35 40 45

Cys Cys Gly Gln Lys Arg Ser Leu Ser Arg Glu Ala Leu Gln Lys Asp  
50 55 60

45 Leu Asp Asp Asn Leu Phe Gly Gln His Leu Ala Lys Lys Ile Ile Leu  
65 70 75 80

Asn Ala Val Phe Gly Phe Ile Asn Asn Pro Lys Pro Lys Lys Pro Leu  
85 90 95

50 Thr Leu Ser Leu His Gly Trp Thr Thr Gly Thr Gly Lys Asn Phe Val Ser  
100 105 110

Lys Ile Ile Ala Glu Asn Ile Tyr Glu Gly Gly Leu Asn Ser Asp Tyr  
115 120 125

55 Val His Leu Phe Val Ala Thr Leu His Phe Pro His Ala Ser Asn Ile  
130 135 140

60 Thr Leu Tyr Lys Asp Gln Leu Leu Trp Ile Arg Gly Asn Val Ser  
145 150 155 160

Ala Cys Ala Arg Ser Ile Phe Ile Phe Asp Glu Met Asp Lys Met His  
165 170 175

5 Ala Gly Leu Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr Asp Leu  
180 185 190

Val Asp Gly Val Ser Tyr Gln Lys Ala Met Phe Ile Phe Leu Ser Asn  
195 200 205

10 Ala Gly Ala Glu Arg Ile Thr Asp Val Ala Leu Asp Phe Trp Arg Ser  
210 215 220

15 Gly Lys Gln Arg Glu Asp Ile Lys Leu Lys Asp Ile Glu His Ala Leu  
225 230 235 240

Ser Val Ser Val Phe Asn Asn Lys Asn Ser Gly Phe Trp His Ser Ser  
245 250 255

20 Leu Ile Asp Arg Asn Leu Ile Asp Tyr Phe Val Pro Phe Leu Pro Leu  
260 265 270

Glu Tyr Lys His Leu Lys Met Cys Ile Arg Val Glu Met Gln Ser Arg  
275 280 285

25 Gly Tyr Glu Ile Asp Glu Asp Ile Val Ser Arg Val Ala Glu Glu Met  
290 295 300

30 Thr Phe Phe Pro Lys Glu Glu Arg Val Phe Ser Asp Lys Gly Cys Lys  
305 310 315 320

Thr Val Phe Thr Lys Leu Asp Tyr Tyr Asp Asp  
325 330

(2) INFORMATION FOR SEQ ID NO: 461:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 461:

Met Leu Lys Cys Ile  
1 5

50 (2) INFORMATION FOR SEQ ID NO: 462:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 462:

Met Ile Leu Thr Leu Leu Ser Val Val Ser Thr Met Ala Ser  
1 5 10



## (2) INFORMATION FOR SEQ ID NO: 463:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 285 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 463:

10 Met Lys Leu His Pro Pro Pro Ser Pro Val Thr Gln Asp His Arg  
1 5 10 15  
15 Ser Lys Ser Ser His Ser Asn Trp Met Pro Arg Met Gly Ala Cys Ser  
20 25 30  
Met Ser Arg Thr Ser Ser Ser Gly Pro Pro Ser Leu Cys Lys Ser Thr  
35 40 45  
20 Ser Gly Arg Ser Cys Thr Arg Pro His Cys Trp Pro Ser Leu Pro Ala  
50 55 60  
Trp Val Ser Val Phe Thr Arg Thr Asn Thr Gly Ser Trp Cys Tyr Pro  
65 70 75 80  
25 Ala Trp Gly Gly Ala Phe Ser Arg Pro Trp Met Ser Ala Gln Ser Met  
85 90 95  
30 Cys Cys Ala Gln Arg Ser Val Leu Gln Val Ala Cys Arg Leu Leu Asp  
100 105 110  
Ala Leu Gln Phe Leu His Gln Asn Gln Tyr Val His Gly Asn Val Thr  
115 120 125  
35 Ala Gln Asn Ile Phe Val Asp Pro Gln Asp Gln Val Thr Leu  
130 135 140  
Ala Gly Tyr Gly Phe Ala Phe Arg Tyr Cys Pro Ser Gly Lys His Val  
145 150 155 160  
40 Ala Tyr Val Gln Gly Ser Arg Ser Pro His Gln Gly Asp Leu Gln Phe  
165 170 175  
45 Ile Ser Met Asp Leu His Lys Gly Cys Gly Pro Ser Arg Arg Xaa Asp  
180 185 190  
Leu Gln Ser Leu Gly Tyr Cys Met Leu Lys Trp Leu Tyr Gly Phe Leu  
195 200 205  
50 Pro Trp Thr Asn Cys Leu Pro Xaa Xaa Gln Asp Ile Met Lys Gln Lys  
210 215 220  
Gln Lys Phe Val Asp Lys Pro Gly Pro Phe Val Gly Pro Cys Gly His  
225 230 235 240  
55 Trp Ile Arg Pro Ser Gln Thr Leu Gln Lys Tyr Leu Lys Val Val Met  
245 250 255  
60 Ala Leu Thr Tyr Gln Gln Lys Pro Pro Tyr Ala Met Leu Arg Asn  
260 265 270

Leu Gln Ala Leu Leu Gln Asp Leu Arg Val Ser Pro Tyr  
275 280 285

## (2) INFORMATION FOR SEQ ID NO: 464:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 464:

15 Met Thr Ser Pro Pro His Gln Gly Trp Gln Gln Arg Gly Cys Gly  
1 5 10 15  
Gln Ser Gln Val Pro Leu Ala Leu Ser Arg Val Phe Ser Thr Ser His  
20 25 30  
Tyr Cys Leu Leu Leu Val Ala Asn Gln Ser Ile Phe Phe Pro Cys Leu  
35 40 45  
25 Trp Ala Val Gln Arg Leu Leu Gly Val Arg Cys Thr Cys Pro Leu Ser  
50 55 60  
Trp Gly Lys Arg Ile Ile Ser Gln His Cys Ser Ala Gln Ser Ser Xaa  
65 70 75 80

## (2) INFORMATION FOR SEQ ID NO: 465:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 465:

45 Met His Thr Trp Tyr Asn Asp Arg Arg Gln Asn Cys His Cys Leu Leu  
1 5 10 15  
Phe Phe Leu Ile Tyr Leu Arg Lys Ile Tyr Gln Val Val Pro His Val  
20 25 30  
50 Pro Leu Leu Val Lys Cys Arg Gly Arg Leu Lys Gly Val Asn Ile  
35 40 45

## (2) INFORMATION FOR SEQ ID NO: 466:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 466:

55  
60

Met Glu Leu Val Leu Val Phe Leu Cys Ser Leu Leu Ala Pro Met Val  
1 5 10 15

5 Leu Ala Ser Ala Ala Glu Lys Glu Lys Glu Met Asp Pro Phe His Tyr  
20 25 30

Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu Val Phe Ala Val Val Leu  
35 40 45

10 Phe Ser Val Gly Ile Leu Leu Ile Leu Ser Arg Cys Lys Cys Ser  
50 55 60

15 Phe Asn Gln Lys Pro Arg Ala Pro Gly Asp Glu Glu Ala Gln Val Glu  
65 70 75 80

Asn Leu Ile Thr Ala Asn Ala Thr Glu Pro Gln Lys Ala Glu Asn Xaa  
85 90 95

20

25 (2) INFORMATION FOR SEQ ID NO: 467:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 399 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 467:

30 Met Ala Ser Gly Ala Asp Ser Lys Gly Asp Asp Leu Ser Thr Ala Ile  
1 5 10 15

Leu Lys Gln Lys Asn Arg Pro Asn Arg Leu Ile Val Asp Glu Ala Ile  
20 25 30

35 Asn Glu Asp Asn Ser Val Val Ser Leu Ser Gln Pro Lys Met Asp Glu  
35 40 45

Leu Gln Leu Phe Arg Gly Asp Thr Val Leu Leu Lys Gly Lys Arg  
50 55 60

45 Arg Glu Ala Val Cys Ile Val Leu Ser Asp Asp Thr Cys Ser Asp Glu  
65 70 75 80

Lys Ile Arg Met Asn Arg Val Val Arg Asn Asn Leu Arg Val Arg Leu  
85 90 95

50 Gly Asp Val Ile Ser Ile Gln Pro Cys Pro Asp Val Lys Tyr Gly Lys  
100 105 110

Arg Ile His Val Leu Pro Ile Asp Asp Thr Val Glu Gly Ile Thr Gly  
115 120 125

55 Asn Leu Phe Glu Val Tyr Leu Lys Pro Tyr Phe Leu Glu Ala Tyr Arg  
130 135 140

60 Pro Ile Arg Lys Gly Asp Ile Phe Leu Val Arg Gly Met Arg Ala

145 150 155 160

Val Glu Phe Lys Val Val Glu Thr Asp Pro Ser Pro Tyr Cys Ile Val  
165 170 175

5 Ala Pro Asp Thr Val Ile His Cys Glu Gly Glu Pro Ile Lys Arg Glu  
180 185 190

10 Asp Glu Glu Ser Leu Asn Glu Val Gly Tyr Asp Asp Ile Gly Gly  
195 200 205

Cys Arg Lys Gln Leu Ala Gln Ile Lys Glu Met Val Glu Leu Pro Leu  
210 215 220

15 Arg His Pro Ala Leu Phe Lys Ala Ile Gly Val Lys Pro Pro Arg Gly  
225 230 235 240

Ile Leu Leu Tyr Gly Pro Gly Thr Gly Lys Thr Leu Ile Ala Arg  
245 250 255

20 Ala Val Ala Asn Glu Thr Gly Ala Phe Phe Leu Ile Asn Gly Pro  
260 265 270

Glu Ile Met Ser Lys Leu Ala Gly Glu Ser Glu Ser Asn Leu Arg Lys  
275 280 285

25 Ala Phe Glu Glu Ala Glu Lys Asn Ala Pro Ala Ile Ile Phe Ile Asp  
290 295 300

30 Glu Leu Asp Ala Ile Ala Pro Lys Arg Glu Lys Thr His Gly Glu Val  
305 310 315 320

Glu Arg Arg Ile Val Ser Gln Leu Leu Thr Leu Met Asp Gly Leu Lys  
325 330 335

35 Gln Arg Ala His Val Ile Val Met Ala Ala Thr Asn Arg Pro Asn Ser  
340 345 350

Ile Asp Pro Ala Leu Arg Arg Phe Gly Arg Phe Asp Arg Glu Val Asp  
355 360 365

40 Ile Gly Ile Pro Asp Ala Thr Gly Arg Leu Glu Ile Leu Gln Ile His  
370 375 380

45 Thr Lys Asn Met Lys Leu Ala Asp Asp Val Asp Leu Glu Gln Xaa  
385 390 395

50 (2) INFORMATION FOR SEQ ID NO: 468:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 468:

55 Leu  
1

60

## (2) INFORMATION FOR SEQ ID NO: 469:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 273 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 469:

10 Met Ala Ala Pro Lys Gly Ser Leu Trp Val Arg Thr Thr Gln Leu Gly Leu  
1 5 10 15  
15 Pro Pro Leu Leu Leu Thr Met Ala Leu Ala Gly Gly Ser Gly Thr  
20 25 30  
Ala Ser Ala Gln Ala Phe Asp Ser Val Leu Gly Asp Thr Ala Ser Cys  
35 40 45  
20 His Arg Ala Cys Gln Leu Thr Tyr Pro Leu His Thr Tyr Pro Lys Gln  
50 55 60  
25 Gln Gln Leu Tyr Ala Cys Gln Arg Gly Cys Arg Leu Phe Ser Ile Cys  
65 70 75 80  
Gln Phe Val Asp Asp Gly Ile Asp Leu Asn Arg Thr Lys Leu Gln Cys  
85 90 95  
30 Gln Ser Ala Cys Thr Gln Ala Tyr Ser Gln Ser Asp Gln Gln Tyr Ala  
100 105 110  
Cys His Leu Gly Cys Gln Asn Gln Leu Pro Phe Ala Gln Leu Arg Gln  
115 120 125  
35 Gln Gln Leu Met Ser Leu Met Pro Lys Met His Leu Leu Phe Pro Leu  
130 135 140  
Thr Leu Val Arg Ser Phe Trp Ser Asp Met Met Asp Ser Ala Gln Ser  
145 150 155 160  
40 Phe Ile Thr Ser Ser Trp Thr Phe Tyr Leu Gln Ala Asp Asp Gly Lys  
165 170 175  
45 Ile Val Ile Phe Xaa Ser Lys Pro Arg Asn Pro Arg Tyr Ala Pro His  
180 185 190  
Leu Gln Pro Gly Ala Leu Pro Asn Leu Xaa Xaa Xaa Ser Leu Ser Lys  
195 200 205  
50 Met Ser Xaa Xaa Ser Xaa Met Arg Asn Ser Gln Ala His Arg Asn Phe  
210 215 220  
Leu Gln Asp Gly Gln Ser Asp Gly Phe Leu Arg Cys Leu Ser Leu Asn  
225 230 235 240  
55 Ser Gly Trp Ile Leu Thr Thr Thr Leu Val Leu Ser Val Met Val Leu  
245 250 255  
60 Leu Trp Ile Cys Cys Ala Thr Cys Cys Tyr Thr Leu Leu Asp Ala Val  
260 265 270

Xaa

## (2) INFORMATION FOR SEQ ID NO: 470:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 470:

15 Met Met Val Leu Ser Leu Gly Ile Ile Leu Ala Ser Ala Ser Phe Ser  
1 5 10 15  
20 Pro Asn Phe Thr Gln Val Thr Ser Thr Leu Leu Asn Ser Ala Tyr Pro  
25 30  
Phe Ile Gly Pro Phe Phe Ile Ile Ser Gly Ser Leu Ser Ile Ala  
35 40 45  
25 Thr Gln Lys Arg Leu Thr Lys Leu Leu Val His Ser Ser Leu Val Gly  
50 55 60  
30 Ser Ile Leu Ser Ala Leu Ser Ala Leu Val Gly Phe Ile Ile Leu Ser  
65 70 75 80  
Val Lys Gln Ala Thr Leu Asn Pro Ala Ser Leu Gln Cys Gln Leu Asp  
85 90 95  
35 Lys Asn Asn Ile Pro Thr Arg Ser Tyr Val Ser Tyr Phe Tyr His Asp  
100 105 110  
Ser Leu Tyr Thr Thr Asp Cys Tyr Thr Ala Lys Ala Ser Leu Ala Gly  
115 120 125  
40 Xaa Leu Ser Leu Met Leu Ile Cys Thr Leu Leu Gln Phe Cys Leu Ala  
130 135 140  
Val Leu Thr Ala Val Leu Arg Trp Lys Gln Ala Tyr Ser Asp Phe Pro  
145 150 155 160  
45 Gly Ser Val Leu Phe Leu Pro His Ser Tyr Ile Gly Asn Ser Gly Met  
165 170 175  
50 Ser Ser Lys Met Thr His Asp Cys Gly Tyr Gln Gln Leu Leu Thr Ser  
180 185 190

## (2) INFORMATION FOR SEQ ID NO: 471:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 234 amino acids

(B) TYPE: amino acid

60

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 471:

5 Met Arg Lys Thr Arg Leu Trp Gly Leu Leu Trp Met Leu Phe Val Ser 15  
1 5 10  
Glu Leu Arg Ala Thr Lys Leu Thr Glu Glu Lys Tyr Glu Leu Lys 20  
20 25 30  
10 Glu Gly Gln Thr Leu Asp Val Lys Cys Asp Tyr Thr Leu Glu Lys Phe 45  
35 40 45  
15 Ala Ser Ser Gln Lys Ala Trp Gln Ile Ile Arg Asp Gly Glu Met Pro 60  
50 60  
Lys Thr Leu Ala Cys Thr Glu Arg Pro Ser Lys Asn Ser His Pro Val 75  
65 80  
20 Gln Val Gly Arg Ile Ile Leu Glu Asp Tyr His Asp His Gly Leu Leu 95  
85 90  
Arg Val Arg Met Val Asn Leu Gln Val Glu Asp Ser Gly Leu Tyr Gln 105  
100 110  
25 Cys Val Ile Tyr Gln Pro Pro Lys Glu Pro His Met Leu Phe Asp Arg 125  
115 120  
Ile Arg Leu Val Val Thr Lys Gly Phe Ser Gly Thr Pro Gly Ser Asn 135  
130 140  
30 Glu Asn Ser Thr Gln Asn Val Tyr Lys Ile Pro Pro Thr Thr Thr Lys 155  
145 160  
Ala Leu Cys Pro Leu Tyr Thr Ser Pro Arg Thr Val Thr Gln Ala Pro 165  
170 175  
35 Pro Lys Ser Thr Ala Asp Val Ser Thr Pro Asp Ser Glu Ile Asn Leu 185  
180 190  
40 Thr Asn Val Thr Asp Ile Ile Arg Val Pro Val Phe Asn Ile Val Ile 195  
200 205  
Leu Leu Ala Gly Gly Phe Leu Ser Lys Ser Leu Val Phe Ser Val Leu 210  
215 220  
45 Phe Ala Val Thr Leu Arg Ser Phe Val Pro 225  
230  
50 (2) INFORMATION FOR SEQ ID NO: 472:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 105 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 472:  
Met Leu His Ile Leu Pro Leu Lys Ser Tyr Asp Phe Pro His Phe Ser 15  
1 5 10

Leu Met Gly Arg Tyr Arg Cys Ala Ser Leu Leu Phe Cys Phe Leu Leu 20  
25 30  
5 Leu Phe Phe Phe Cys Ser Val Leu Trp Thr Phe Ser Asp Met His 35  
40 45  
Arg Ser Gly Glu Asp Gly Pro Trp Thr Pro Cys Val His His Leu Ala 50  
55 60  
10 Ala Ser Leu Ile Ser Tyr Gly Gln Pro Gly Phe Ile Cys Ile Ser Leu 65  
70 75 80  
15 Phe Ser Pro Val Leu Phe Ile Glu Asn Pro Arg His Tyr Ala Asn Ala 85  
90 95  
Thr Val Thr Thr Leu Gly Asp Trp Xaa 100  
105  
20 (2) INFORMATION FOR SEQ ID NO: 473:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 473:  
Met Val Phe Leu Lys Tyr Arg Phe Leu Phe Phe Leu Val Phe Leu Ala 1  
5 10 15  
30 Asn Cys Ile Tyr Ser Leu His Tyr Lys Pro Ser Leu Met Tyr Pro Lys 20  
25 30  
35 (2) INFORMATION FOR SEQ ID NO: 474:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 571 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 474:  
Met Ala Leu Ser Arg Gly Leu Pro Arg Glu Leu Ala Glu Ala Val Ala 1  
5 10 15  
Gly Gly Arg Val Leu Val Val Gly Ala Gly Gly Ile Gly Cys Glu Leu 20  
25 30  
55 Leu Lys Asn Leu Val Leu Thr Gly Phe Ser His Ile Asp Leu Ile Asp 35  
40 45  
Leu Asp Thr Ile Asp Val Ser Asn Leu Asn Arg Gln Phe Leu Phe Gln 50  
55 60

Lys Lys His Val Gly Arg Ser Lys Ala Gln Val Ala Lys Glu Ser Val  
65 70 75 80  
5 Leu Gln Phe Tyr Pro Lys Ala Asn Ile Val Ala Tyr His Asp Ser Ile  
85 90 95  
Met Asn Pro Asp Tyr Asn Val Glu Phe Arg Arg Gln Phe Ile Leu Val  
100 105 110  
10 Met Asn Ala Leu Asp Asn Arg Ala Ala Arg Asn His Val Asn Arg Met  
115 120 125  
15 Cys Leu Ala Ala Asp Val Pro Leu Ile Glu Ser Gly Thr Ala Gly Tyr  
130 135 140  
Leu Gly Gln Val Thr Thr Ile Lys Lys Gly Val Thr Glu Cys Tyr Glu  
145 150 155 160  
20 Cys His Pro Lys Pro Thr Gln Arg Thr Phe Pro Gly Cys Thr Ile Arg  
165 170 175  
Asn Thr Pro Ser Glu Pro Ile His Cys Ile Val Trp Ala Lys Tyr Leu  
180 185 190  
25 Phe Asn Gln Leu Phe Gly Glu Glu Asp Ala Asp Gln Glu Val Ser Pro  
195 200 205  
30 Asp Arg Ala Asp Pro Glu Ala Ala Trp Glu Pro Thr Glu Ala Glu Ala  
210 215 220  
Arg Ala Arg Ala Ser Asn Glu Asp Gly Asp Ile Lys Arg Ile Ser Thr  
225 230 235 240  
35 Lys Glu Trp Ala Lys Ser Thr Gly Tyr Asp Pro Val Lys Leu Thr  
245 250 255  
Lys Leu Phe Lys Asp Asp Ile Arg Tyr Leu Leu Thr Met Asp Lys Leu  
260 265 270  
40 Trp Arg Lys Arg Lys Pro Pro Val Pro Leu Asp Trp Ala Glu Val Gln  
275 280 285  
Ser Gln Gly Glu Glu Thr Asn Ala Ser Asp Gln Gln Asn Glu Pro Gln  
290 295 300  
45 Leu Gly Leu Lys Asp Gln Gln Val Leu Asp Val Lys Ser Tyr Ala Arg  
305 310 315 320  
50 Leu Phe Ser Lys Ser Ile Glu Thr Leu Arg Val His Leu Ala Glu Lys  
325 330 335  
Gly Asp Gly Ala Glu Leu Ile Trp Asp Lys Asp Asp Pro Ser Ala Met  
340 345 350  
55 Asp Phe Val Thr Ser Ala Ala Asn Leu Arg Met His Ile Phe Ser Met  
355 360 365  
Asn Met Lys Ser Arg Phe Asp Ile Lys Ser Met Ala Gly Asn Ile Ile  
370 375 380 385 390 395 400

Pro Ala Ile Ala Thr Thr Asn Ala Val Ile Ala Gly Leu Ile Val Leu  
385 390 395 400  
5 Glu Gly Leu Lys Ile Leu Ser Gly Lys Ile Asp Gln Cys Arg Thr Ile  
405 410 415  
Phe Leu Asn Lys Gln Pro Asn Pro Arg Lys Lys Leu Leu Val Pro Cys  
420 425 430  
10 Ala Leu Asp Pro Pro Asn Pro Asn Cys Tyr Val Cys Ala Ser Lys Pro  
435 440 445  
15 Glu Val Thr Val Arg Leu Asn Val His Lys Val Thr Val Leu Thr Leu  
450 455 460  
Gln Asp Lys Ile Val Lys Glu Lys Phe Ala Met Val Ala Pro Asp Val  
465 470 475 480  
20 Gln Ile Glu Asp Gly Lys Gly Thr Ile Leu Ile Ser Ser Glu Glu Gly  
485 490 495  
Glu Thr Glu Ala Asn Asn His Lys Lys Leu Ser Glu Phe Gly Ile Arg  
500 505 510  
25 Asn Gly Ser Arg Leu Gln Ala Asp Asp Phe Leu Gln Asp Tyr Thr Leu  
515 520 525  
30 Leu Ile Asn Ile Leu His Ser Glu Asp Leu Gly Lys Asp Val Glu Phe  
530 535 540  
Glu Val Val Gly Asp Ala Pro Glu Lys Val Gly Xaa Lys Gln Ala Glu  
545 550 555 560  
35 Asp Ala Ala Lys Ser Ile Thr Asn Gly Gln Xaa  
565 570 575  
40 (2) INFORMATION FOR SEQ ID NO: 475:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 312 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
45 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 475:  
Met Gln Val Val Thr Cys Leu Thr Arg Asp Ser Tyr Leu Thr His Cys  
1 5 10 15  
50 Phe Leu Gln His Leu Met Val Val Leu Ser Ser Leu Glu Arg Thr Pro  
20 25 30  
Ser Pro Glu Pro Val Asp Lys Asp Phe Tyr Ser Glu Phe Gly Asn Lys  
35 40 45  
55 Thr Thr Gly Lys Met Glu Asn Tyr Glu Leu Ile His Ser Ser Arg Val  
50 55 60  
Lys Phe Thr Tyr Pro Ser Glu Glu Glu Ile Gly Asp Leu Thr Phe Thr  
65 70 75 80 85 90 95 100

Val Ala Gln Lys Met Ala Glu Pro Glu Lys Ala Pro Ala Leu Ser Ile  
85 90 95

5 Leu Leu Tyr Val Gln Ala Phe Gln Val Glu Met Pro Pro Pro Gly Cys  
100 105 110

Cys Arg Gly Pro Leu Arg Pro Lys Thr Leu Leu Thr Ser Ser Glu  
115 120 125

10 Ile Phe Leu Leu Asp Glu Asp Cys Val His Tyr Pro Leu Pro Glu Phe  
130 135 140

Ala Lys Glu Pro Pro Gln Arg Asp Arg Tyr Arg Leu Asp Asp Gly Arg  
145 150 155 160

Arg Val Arg Asp Leu Asp Arg Val Leu Met Gly Tyr Gln Thr Tyr Pro  
165 170 175

20 Gln Pro Ser Pro Ser Ser Met Thr Cys Lys Val Met Thr Ser Trp  
180 185 190

Ala Val Ser Pro Trp Thr Thr Leu Gly Arg Cys Gln Val Ala Arg Leu  
195 200 205

25 Glu Pro Ala Arg Ala Val Lys Ser Ser Gly Arg Cys Leu Ser Pro Val  
210 215 220

Leu Arg Ala Glu Arg Ser Ser Arg Cys Trp Leu Ala Ser Gly Arg  
225 230 235 240

Pro Cys Val Ala Val Ser Cys Leu Ser Ser Pro Ala Ser Pro Gly  
245 250 255

35 His Ser Gln Pro Val Val Ser Ser Leu Thr Pro Thr Gly Ala Gly Gln  
260 265 270

Gln Ala Phe Val Phe Ser Lys Asn Val Leu Ser Ser Leu Trp Tyr Leu  
275 280 285

40 Asn Leu Thr Val Leu Ala Glu Asn Val Asn Met Cys Val Cys Cys Val  
290 295 300

Asn Ser Phe Ser Cys Trp Glu Xaa  
305 310

(2) INFORMATION FOR SEQ ID NO: 476:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 329 amino acids  
(B) TYPE: amino acid  
(C) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 476:

Met Ala Gln His His Leu Trp Ile Leu Leu Cys Leu Gln Thr Trp  
1 5 10 15

60 Pro Glu Ala Ala Gly Lys Asp Ser Glu Ile Phe Thr Val Asn Gly Ile

20 25 30

Leu Gly Glu Ser Val Thr Phe Pro Val Asn Ile Gln Glu Pro Arg Gln  
35 40 45

5 Val Lys Ile Ile Ala Trp Thr Ser Lys Thr Ser Val Ala Tyr Val Thr  
50 55 60

Pro Gly Asp Ser Glu Thr Ala Pro Val Val Thr Val Thr His Arg Asn  
65 70 75 80

10 Tyr Tyr Glu Arg Ile His Ala Leu Gly Pro Asn Tyr Asn Leu Val Ile  
85 90 95

15 Ser Asp Leu Arg Met Glu Asp Ala Gly Asp Tyr Lys Ala Asp Ile Asn  
100 105 110

Thr Gln Ala Asp Pro Tyr Thr Thr Lys Arg Tyr Asn Leu Gln Ile  
115 120 125

20 Tyr Arg Arg Leu Gly Lys Pro Lys Ile Thr Gln Ser Leu Met Ala Ser  
130 135 140

Val Asn Ser Thr Cys Asn Val Thr Leu Thr Cys Ser Val Glu Lys Glu  
145 150 155 160

Glu Lys Asn Val Thr Tyr Asn Trp Ser Pro Leu Gly Glu Gly Asn  
165 170 175

30 Val Leu Gln Ile Phe Gln Thr Pro Glu Asp Gln Glu Leu Thr Tyr Thr  
180 185 190

Cys Thr Ala Gln Asn Pro Val Ser Asn Asn Ser Asp Ser Ile Ser Ala  
195 200 205

35 Arg Gln Leu Cys Ala Asp Ile Ala Met Gly Phe Arg Thr His His Thr  
210 215 220

Gly Leu Leu Ser Val Leu Ala Met Phe Phe Leu Leu Val Leu Ile Leu  
225 230 235 240

Ser Ser Val Phe Leu Phe Arg Leu Phe Lys Arg Arg Gln Asp Ala Ala  
245 250 255

45 Ser Lys Lys Thr Ile Tyr Thr Tyr Ile Met Ala Ser Arg Asn Thr Gln  
260 265 270

Pro Ala Glu Ser Arg Ile Tyr Asp Glu Ile Leu Gln Ser Lys Val Leu  
275 280 285

50 Pro Ser Lys Glu Glu Pro Val Asn Thr Val Tyr Ser Glu Val Gln Phe  
290 295 300

Ala Asp Lys Met Gly Lys Ala Ser Thr Gln Asp Ser Lys Pro Pro Gly  
305 310 315 320

Thr Ser Ser Tyr Glu Ile Val Ile Xaa  
325

60

## (2) INFORMATION FOR SEQ ID NO: 477:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 178 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 477:

10 Met Lys Leu Gln Cys Val Ser Leu Trp Leu Leu Gly Thr Ile Leu Ile  
1 5 10 15  
Leu Cys Ser Val Asp Asn His Gly Leu Arg Arg Cys Leu Ile Ser Thr  
20 25 30  
Asp Met His His Ile Gln Gln Ser Phe Gln Gln Ile Lys Arg Ala Ile  
35 40 45  
Gln Ala Lys Asp Thr Phe Pro Asn Val Thr Ile Leu Ser Thr Leu Gln  
50 55 60  
Thr Leu Gln Ile Ile Lys Pro Leu Asp Val Cys Cys Val Thr Lys Asn  
65 70 75 80  
Leu Leu Ala Phe Tyr Val Asp Arg Val Phe Lys Asp His Gln Gln Pro  
85 90 95  
Asn Pro Lys Ile Leu Arg Lys Ile Ser Ser Ile Ala Asn Ser Phe Leu  
100 105 110  
Tyr Met Gln Lys Thr Leu Arg Gln Cys Gln Gln Arg Gln Cys His  
115 120 125  
Cys Arg Gln Gln Ala Thr Asn Ala Thr Arg Val Ile His Asp Asn Tyr  
130 135 140  
Asp Gln Leu Gln Val His Ala Ala Ile Lys Ser Leu Gly Gln Leu  
145 150 155 160  
Asp Val Phe Leu Ala Trp Ile Asn Lys Asn His Gln Val Met Ser Ser  
165 170 175  
Ala Xaa

45

## (2) INFORMATION FOR SEQ ID NO: 478:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 478:

55 Asp Thr Ala Ile Arg Val Ala Leu Ala Val Ala Val Leu Lys Thr Val  
1 5 10 15  
Ile Leu Gly Leu Leu Cys Leu Leu Leu Cys Gly Gly Gln Gly Lys  
20 25 30

60

Val Ala Gly Arg Gln Ala Val Thr Ser Asp Gln Gln Ser Val Gly Arg  
35 40 45

5 Arg Asp Val Tyr  
50

## (2) INFORMATION FOR SEQ ID NO: 479:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 62 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 479:

10 Met Gln Lys Lys Asn Ser Leu Phe Phe Phe Ala Phe Tyr Tyr Gln  
1 5 10 15  
Asn Lys Thr Asn Ala Pro Gly Gln Gly Ser Met Ile Thr Arg Asn Ile  
20 25 30  
Lys Gln Tyr Phe Leu Pro Phe Leu Phe Cys Cys Val Gln Ala Ser Ile  
35 40 45  
Ala Ile Asn Lys Leu Asn Tyr Leu His Trp Thr His Phe Gln  
50 55 60

30

## (2) INFORMATION FOR SEQ ID NO: 480:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 480:

40 Met Pro Gly Leu Ser Leu Ile Leu Thr Val Thr Leu Leu Ala Val Ser  
1 5 10 15  
Asp Ser Ala Ala Thr Cys Ile Val Ala Lys Gly  
20 25

45

## (2) INFORMATION FOR SEQ ID NO: 481:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 339 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 481:

55 Met Ser Gly Pro Asp Val Gln Thr Pro Ser Ala Ile Gln Ile Cys Arg  
1 5 10 15  
Ile Met Arg Pro Asp Asp Ala Asn Val Ala Gly Asn Val His Gly Gly  
20 25 30

60

645

Thr Ile Leu Lys Met Ile Glu Glu Ala Gly Ala Ile Ile Ser Thr Arg  
35 40 45

5 His Cys Asn Ser Gln Asn Gly Glu Arg Cys Val Ala Ala Leu Ala Arg  
50 55 60

Val Glu Arg Thr Asp Phe Leu Ser Pro Met Cys Ile Gly Glu Val Ala  
65 70 75 80

10 His Val Ser Ala Glu Ile Thr Tyr Thr Ser Lys His Ser Val Glu Val  
85 90 95

15 Gln Val Asn Val Met Ser Glu Asn Ile Leu Thr Gly Ala Lys Lys Leu  
100 105 110

Thr Asn Lys Ala Thr Leu Trp Tyr Val Pro Leu Ser Leu Lys Asn Val  
115 120 125

20 Asp Lys Val Leu Glu Val Pro Val Val Tyr Ser Arg Xaa Glu Gln  
130 135 140

Glu Glu Glu Gly Arg Lys Arg Tyr Glu Ala Gln Lys Leu Glu Arg Met  
145 150 155 160

25 Glu Thr Lys Trp Arg Asn Gly Asp Ile Val Gln Pro Val Leu Asn Pro  
165 170 175

30 Glu Pro Asn Thr Val Ser Tyr Ser Gln Ser Ser Leu Ile His Leu Val  
180 185 190

Gly Pro Ser Asp Cys Thr Leu His Gly Phe Val His Gly Gly Val Thr  
195 200 205

35 Met Lys Leu Met Asp Glu Val Ala Gly Ile Val Ala Ala Arg His Cys  
210 215 220

Lys Thr Asn Ile Val Thr Ala Ser Val Asp Ala Ile Asn Phe His Asp  
225 230 235 240

40 Lys Ile Arg Lys Gly Cys Val Ile Thr Ile Ser Gly Arg Met Thr Phe  
245 250 255

45 Thr Ser Asn Lys Ser Met Glu Ile Glu Val Leu Val Asp Ala Asp Pro  
260 265 270

Val Val Asp Ser Ser Gln Lys Arg Tyr Arg Ala Ala Ser Ala Phe Phe  
275 280 285

50 Thr Tyr Val Ser Leu Ser Gln Glu Gly Arg Ser Leu Pro Val Pro Gln  
290 295 300

Leu Val Pro Glu Thr Glu Asp Glu Lys Lys Arg Phe Glu Glu Gly Lys  
305 310 315 320

55 Gly Arg Tyr Leu Gln Met Lys Ala Lys Xaa Gln Gly His Ala Xaa Xaa  
325 330 335

Gln Pro Xaa

60

646

(2) INFORMATION FOR SEQ ID NO: 482:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 482:

Met Leu Asn Ser Asn Ile Asn Asp Leu Leu Met Val Thr Tyr Leu Ala  
1 5 10 15

15 Asn Leu Thr Gln Ser Gln Ile Ala Leu Asn Glu Lys Leu Val Asn Leu  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 483:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 48 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 483:

Met Arg Glu Thr Ser Ile Arg Val Leu Leu Met Leu Pro Ala Leu Glu  
1 5 10 15

35 Ser Thr Ser Gly Leu Ser Ala Phe Met Gly Leu Gly Thr Arg Ile Gly  
20 25 30

Cys Phe Lys Thr Ile Thr Cys Trp Pro Thr Ser Leu Thr Gln Arg Xaa  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 484:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 38 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 484:

Met Tyr Met Tyr Ser Leu Asn Val Phe Leu Ser Phe Ile Phe Leu Ala  
1 5 10 15

55 Leu Val Phe Lys Cys Val His Val Cys Gln Gly Ala Asn Ala Phe Leu  
20 25 30

Phe Leu Lys Leu Val Phe

35



- (2) INFORMATION FOR SEQ ID NO: 485:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 61 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 485:
- Met Gly Leu Arg Leu Ile Cys Leu Glu Leu Thr Met Val Lys Ala Leu  
1 5 10 15  
Val Cys Glu Met Phe Leu Phe Leu Met Thr Gln Lys Leu Ile Thr  
20 25 30  
Gln Glu Cys Thr Glu Lys Phe Ala Lys Leu Leu Val Gln Leu Ile Ser  
35 40 45  
Leu Val Phe Ala Trp Glu Phe Phe Ser Glu Asp Thr Pro  
50 55 60
- (2) INFORMATION FOR SEQ ID NO: 486:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 346 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 486:
- Met Leu Ala Ala Arg Leu Val Cys Leu Arg Thr Leu Pro Ser Arg Val  
1 5 10 15  
Phe His Pro Ala Phe Thr Lys Ala Ser Pro Val Val Lys Asn Ser Ile  
20 25 30  
Thr Lys Asn Gln Trp Leu Leu Thr Pro Ser Arg Glu Tyr Ala Thr Lys  
35 40 45  
Thr Arg Ile Gly Ile Arg Arg Gly Arg Thr Gly Gln Glu Leu Lys Glu  
50 55 60  
Ala Ala Leu Glu Pro Ser Met Glu Lys Ile Phe Lys Ile Asp Gln Met  
65 70 75 80  
Gly Arg Trp Phe Val Ala Gly Gly Ala Ala Val Gly Leu Gly Ala Leu  
85 90 95  
Cys Tyr Tyr Gly Leu Gly Leu Ser Asn Glu Ile Gly Ala Ile Glu Lys  
100 105 110  
Ala Val Ile Trp Pro Gln Tyr Val Lys Asp Arg Ile His Ser Thr Tyr  
115 120 125  
Met Tyr Leu Ala Gly Ser Ile Gly Leu Thr Ala Leu Ser Ala Ile Ala  
130 135 140

- Ile Ser Arg Thr Pro Val Leu Met Asn Phe Met Met Arg Gly Ser Trp  
145 150 155 160  
Val Thr Ile Gly Val Thr Phe Ala Met Val Gly Ala Gly Met Leu  
165 170 175  
Val Arg Ser Ile Pro Tyr Asp Gln Ser Pro Gly Pro Lys His Leu Ala  
180 185 190  
Trp Leu Leu His Ser Gly Val Met Gly Ala Val Val Ala Pro Leu Thr  
195 200 205  
Ile Leu Gly Gly Pro Leu Leu Ile Arg Ala Ala Trp Tyr Thr Ala Gly  
210 215 220  
Ile Val Gly Gly Leu Ser Thr Val Ala Met Cys Ala Pro Ser Glu Lys  
225 230 235 240  
Phe Leu Asn Met Gly Ala Pro Leu Gly Val Gly Leu Gly Leu Val Phe  
245 250 255  
Val Ser Ser Leu Gly Ser Met Phe Leu Pro Pro Thr Thr Val Ala Gly  
260 265 270  
Ala Thr Leu Tyr Ser Val Ala Met Tyr Gly Gly Leu Val Leu Phe Ser  
275 280 285  
Met Phe Leu Leu Tyr Asp Thr Gln Lys Val Ile Lys Arg Ala Glu Val  
290 295 300  
Ser Pro Met Tyr Gly Val Gln Lys Tyr Asp Pro Ile Asn Ser Met Leu  
305 310 315 320  
Ser Ile Tyr Met Asp Thr Leu Asn Ile Phe Met Arg Val Ala Thr Met  
325 330 335  
Leu Ala Thr Gly Gly Asn Arg Lys Lys Xaa  
340 345
- (2) INFORMATION FOR SEQ ID NO: 487:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 237 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 487:
- Met Glu Glu Val Leu Leu Leu Gly Leu Lys Asp Arg Glu Gly Tyr Thr  
1 5 10 15  
Ser Phe Trp Asn Asp Cys Ile Ser Ser Gly Leu Arg Gly Cys Met Leu  
20 25 30  
Ile Glu Leu Ala Leu Arg Gly Arg Leu Gln Leu Ala Cys Gly Met  
35 40 45  
Arg Arg Lys Ser Leu Leu Thr Arg Lys Val Ile Cys Lys Ser Asp Ala  
50 55 60

649

Pro Thr Gly Asp Val Leu Leu Asp Glu Ala Leu Lys His Val Lys Glu  
65 70 75 80

5 Thr Gln Pro Pro Glu Thr Val Gln Asn Trp Ile Glu Leu Leu Ser Gly  
85 90 95

Glu Thr Trp Asn Pro Leu Lys Leu His Tyr Gln Leu Arg Asn Val Arg  
100 105 110

10 Glu Arg Leu Ala Lys Asn Leu Val Glu Lys Gly Val Leu Thr Thr Glu  
115 120 125

Lys Gln Asn Phe Leu Leu Phe Asp Met Thr Thr His Pro Leu Thr Asn  
130 135 140

Asn Asn Ile Lys Gln Arg Leu Ile Lys Lys Val Gln Glu Ala Val Leu  
145 150 155

20 Asp Lys Trp Val Asn Asp Pro His Arg Met Asp Arg Arg Leu Leu Ala  
165 170 175

Leu Ile Tyr Leu Ala Ala Ser Asp Val Leu Glu Asn Ala Phe Ala  
180 185 190

25 Pro Leu Leu Asp Glu Gln Tyr Asp Leu Ala Thr Lys Arg Val Arg Gln  
195 200 205

30 Leu Leu Asp Leu Asp Pro Glu Val Glu Cys Leu Lys Ala Asn Thr Asn  
210 215 220

Glu Val Leu Trp Ala Val Val Ala Phe Thr Lys Xaa  
225 230 235

35

(2) INFORMATION FOR SEQ ID NO: 488:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 200 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 488:

45 Met Ala Gln Arg Met Val Trp Val Asp Leu Glu Met Thr Gly Leu Asp  
1 5 10 15

Ile Glu Lys Asp Gln Ile Ile Glu Met Ala Cys Leu Ile Thr Asp Ser  
20 25 30

50 Asp Leu Asn Ile Leu Ala Glu Gly Pro Asn Leu Ile Ile Lys Gln Pro  
35 40 45

55 Asp Glu Leu Leu Asp Ser Met Ser Asp Trp Cys Lys Glu His His Gly  
50 55 60

Lys Ser Gly Leu Thr Lys Ala Val Lys Glu Ser Thr Ile Thr Leu Gln  
65 70 75 80

60 Gln Ala Glu Tyr Glu Phe Leu Ser Phe Val Arg Gln Gln Thr Pro Pro

650

85 90 95

Gly Leu Cys Pro Leu Ala Gly Asn Ser Val His Glu Asp Lys Lys Phe  
100 105 110

5 Leu Asp Lys Tyr Met Pro Gln Phe Met Lys His Leu His Tyr Arg Ile  
115 120 125

10 Ile Asp Val Ser Thr Val Lys Glu Leu Cys Arg Arg Trp Tyr Pro Glu  
130 135 140

Glu Tyr Glu Phe Ala Pro Lys Lys Ala Ala Ser His Arg Ala Leu Asp  
145 150 155

15 Asp Ile Ser Glu Ser Ile Lys Glu Leu Gln Phe Tyr Arg Asn Ile  
165 170 175

Phe Lys Lys Lys Ile Asp Glu Lys Lys Arg Lys Ile Ile Glu Asn Gly  
180 185 190

20 Glu Asn Glu Lys Thr Val Ser Xaa  
195 200

(2) INFORMATION FOR SEQ ID NO: 489:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 351 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 489:

35 Met Ala Thr Thr Ala Ala Pro Ala Gly Gly Ala Arg Asn Gly Ala Gly  
1 5 10 15

Pro Glu Trp Gly Gly Phe Glu Glu Asn Ile Gln Gly Gly Gly Ser Ala  
20 25 30

40 Val Ile Asp Met Glu Asn Met Asp Asp Thr Ser Gly Ser Ser Phe Glu  
35 40 45

Asp Met Gly Glu Leu His Gln Arg Leu Arg Glu Glu Val Asp Ala  
50 55 60

45 Asp Ala Ala Asp Ala Ala Ala Glu Glu Asp Gly Glu Phe Leu  
65 70 75 80

Gly Met Lys Gly Phe Lys Gly Gln Leu Ser Arg Gln Val Ala Asp Gln  
85 90 95

50 Met Trp Gln Ala Gly Lys Arg Gln Ala Ser Arg Ala Phe Ser Leu Tyr  
100 105 110

55 Ala Asn Ile Asp Ile Leu Arg Pro Tyr Phe Asp Val Glu Pro Ala Gln  
115 120 125

Val Arg Thr Gly Leu Leu Glu Ser Met Ile Pro Ile Lys Met Val Asn  
130 135 140

60

Phe Pro Gln Lys Ile Ala Gly Gln Leu Tyr Gly Pro Leu Met Leu Val  
145 150 155 160  
Phe Thr Leu Val Ala Ile Leu Leu His Gly Met Lys Thr Ser Asp Thr  
165 170 175  
Ile Ile Arg Gln Gly Thr Leu Met Gly Thr Ala Ile Gly Thr Cys Phe  
180 185 190  
10 Gly Tyr Trp Leu Gly Val Ser Ser Phe Ile Tyr Phe Leu Ala Tyr Leu  
195 200 205  
Cys Asn Ala Gln Ile Thr Met Leu Gln Met Leu Ala Leu Leu Gly Tyr  
210 215 220  
15 Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr Tyr Asn Ile His  
225 230 235 240  
Leu His Ala Leu Phe Tyr Leu Phe Trp Leu Leu Val Gly Gly Leu Ser  
245 250 255  
20 Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr Val Gly Pro Thr  
260 265 270  
25 Gln Arg Leu Leu Leu Cys Gly Thr Leu Ala Ala Leu His Met Leu Phe  
275 280 285  
Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Val Gln Gly Ile Leu  
290 295 300  
30 Asp Thr Leu Gln Gly Pro Asn Ile Pro Pro Ile Gln Arg Val Pro Arg  
305 310 315 320  
Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro Thr Thr Val Leu  
325 330 335  
Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln Ser His Xaa  
340 345 350  
40 (2) INFORMATION FOR SEQ ID NO: 490:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 265 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 490:  
Met Arg Gly Ser Arg Gly Gly Tyr Ala Gly Gln Met Ala Ala Ser Gly  
1 5 10 15  
Glu Ser Gly Thr Ser Gly Gly Gly Ser Thr Gln Ala Ala Phe Met  
20 25 30  
Thr Phe Tyr Ser Gln Val Lys Gln Ile Gln Lys Arg Asp Ser Val Leu  
35 40 45  
Thr Ser Lys Asn Gln Ile Gln Arg Leu Thr Arg Pro Gly Ser Ser Tyr  
50 55 60

Phe Asn Leu Asn Pro Phe Gln Val Leu Gln Ile Asp Pro Gln Val Thr  
65 70 75 80  
5 Asp Gln Gln Ile Lys Lys Arg Phe Arg Gln Leu Ser Ile Leu Val His  
85 90 95  
Pro Asp Lys Asn Gln Asp Asp Ala Asp Arg Ala Gln Lys Ala Phe Gln  
100 105 110  
10 Ala Val Asp Lys Ala Tyr Lys Leu Leu Leu Asp Gln Gln Lys Lys  
115 120 125  
15 Arg Ala Leu Asp Val Ile Gln Ala Gly Lys Gln Tyr Val Gln His Thr  
130 135 140  
Val Lys Gln Arg Lys Lys Gln Leu Lys Lys Gln Gly Lys Pro Thr Ile  
145 150 155 160  
20 Val Gln Gln Asp Asp Pro Gln Leu Phe Lys Gln Ala Val Tyr Lys Gln  
165 170 175  
Thr Met Lys Leu Phe Ala Gln Leu Gln Ile Lys Arg Lys Gln Arg Gln  
180 185 190  
25 Ala Lys Gln Met His Gln Arg Lys Arg Gln Arg Gln Gln Gln Ile Gln  
195 200 205  
30 Ala Gln Gln Lys Ala Lys Arg Gln Arg Gln Trp Gln Lys Asn Phe Gln  
210 215 220  
Gln Ser Arg Asp Gly Arg Val Asp Ser Trp Arg Asn Phe Gln Ala Asn  
225 230 235 240  
35 Thr Lys Gly Lys Lys Gln Lys Lys Asn Arg Thr Phe Leu Arg Pro Pro  
245 250 255  
Lys Val Lys Met Gln Gln Arg Gln Xaa  
260 265  
40 (2) INFORMATION FOR SEQ ID NO: 491:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 491:  
Asp Ser Met Pro Thr Cys Pro Leu Xaa Ala Ser Leu Gln Cys Gly Pro  
1 5 10 15  
Leu Leu Pro Val Arg Leu Cys Cys Leu  
20 25  
55 (2) INFORMATION FOR SEQ ID NO: 492:  
60

653

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 159 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 492:

Met Asn Glu Tyr Arg Val Pro Glu Leu Asn Val Gln Asn Gly Val Leu  
1 5 10 15

Lys Ser Leu Ser Phe Leu Phe Glu Tyr Ile Gly Glu Met Gly Lys Asp  
20 25 30

Tyr Ile Tyr Ala Val Thr Pro Leu Leu Glu Asp Ala Leu Met Asp Arg  
35 40 45

Asp Leu Val His Arg Gln Thr Ala Ser Ala Val Val Gln His Met Ser  
50 55 60

Leu Gly Val Tyr Gly Phe Gly Cys Glu Asp Ser Leu Asn His Leu Leu  
65 70 75 80

Asn Tyr Val Trp Pro Asn Val Phe Glu Thr Ser Pro His Val Ile Gln  
85 90 95

Ala Val Met Gly Ala Leu Glu Gly Leu Arg Val Ala Ile Gly Pro Cys  
100 105 110

Arg Met Leu Gln Tyr Cys Leu Gln Gly Leu Phe His Pro Ala Arg Lys  
115 120 125

Val Arg Asp Val Tyr Trp Lys Ile Tyr Asn Ser Ile Tyr Ile Gly Ser  
130 135 140

Gln Asp Ala Leu Ile Ala His Tyr Pro Arg Ile Tyr Gln Arg Xaa  
145 150 155

(2) INFORMATION FOR SEQ ID NO: 493:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 279 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 493:

Met Ile Ser Asp Asn Ser Ala Glu Asn Ile Ala Leu Val Thr Ser Met  
1 5 10 15

Tyr Asp Gly Leu Leu Gln Ala Gly Ala Arg Leu Cys Pro Thr Val Gln  
20 25 30

Leu Glu Asp Ile Arg Asn Leu Gln Asp Leu Thr Pro Leu Lys Leu Ala  
35 40 45

Ala Lys Glu Gly Lys Ile Glu Ile Phe Arg His Ile Leu Gln Arg Glu  
50 55 60

Phe Ser Gly Leu Ser His Leu Ser Arg Lys Phe Thr Glu Trp Cys Tyr  
65 70 75 80

WO 98/39448

654

Gly Pro Val Arg Val Ser Leu Tyr Asp Leu Ala Ser Val Asp Ser Cys  
85 90 95

Glu Glu Asn Ser Val Leu Glu Ile Ile Ala Phe His Cys Lys Ser Pro  
100 105 110

His Arg His Arg Met Val Val Leu Glu Pro Leu Asn Lys Leu Leu Gln  
115 120 125

Ala Lys Trp Asp Leu Leu Ile Pro Lys Phe Leu Asn Phe Leu Cys  
130 135 140

Asn Leu Ile Tyr Met Phe Ile Phe Thr Ala Val Ala Tyr His Gln Pro  
145 150 155 160

Thr Leu Lys Lys Gln Ala Ala Pro His Leu Lys Ala Glu Val Gly Asn  
165 170 175

Ser Met Leu Leu Thr Gly His Ile Leu Ile Leu Leu Gly Gly Ile Tyr  
180 185 190

Leu Leu Val Gly Gln Leu Trp Tyr Phe Trp Arg Arg His Val Phe Ile  
195 200 205

Trp Ile Ser Phe Ile Asp Ser Tyr Phe Glu Ile Leu Phe Leu Phe Gln  
210 215 220

Ala Leu Leu Thr Val Val Ser Gln Val Leu Cys Phe Leu Xaa Ile Glu  
225 230 235 240

Trp Tyr Leu Pro Leu Leu Val Ser Ala Leu Val Leu Gly Trp Leu Asn  
245 250 255

Leu Leu Tyr Tyr Thr Arg Gly Phe Gln His Thr Gly Ile Tyr Ser Val  
260 265 270

Met Ile Gln Lys Pro Trp Xaa  
275

(2) INFORMATION FOR SEQ ID NO: 494:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 193 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 494:

Met Ile Arg Cys Gly Leu Ala Cys Glu Arg Cys Arg Trp Ile Leu Pro  
1 5 10 15

Leu Leu Leu Ser Ala Ile Ala Phe Asp Ile Ile Ala Leu Ala Gly  
20 25 30

Arg Gly Trp Leu Gln Ser Ser Asp His Gly Gln Thr Ser Ser Leu Trp  
35 40 45

Trp Lys Cys Ser Gln Glu Gly Gly Ser Gly Ser Tyr Glu Glu Gly  
50 55 60

655

50 55 60  
Cys Gln Ser Leu Met Glu Tyr Ala Trp Gly Arg Ala Ala Ala Met  
65 70 75 80  
Leu Phe Cys Gly Phe Ile Ile Leu Val Ile Cys Phe Ile Leu Ser Phe  
85 90 95  
Phe Ala Leu Cys Gly Pro Gln Met Leu Val Phe Leu Arg Val Ile Gly  
100 105 110  
Gly Leu Leu Ala Leu Ala Val Phe Gln Ile Ile Ser Leu Val Ile  
115 120 125  
Tyr Pro Val Lys Tyr Thr Gln Thr Phe Thr Leu His Ala Asn Xaa Ala  
130 135 140  
Val Thr Tyr Ile Tyr Asn Trp Ala Tyr Gly Phe Gly Trp Ala Ala Thr  
145 150 155 160  
Ile Ile Leu Ile Gly Cys Ala Phe Phe Cys Cys Leu Pro Asn Tyr  
165 170 175  
Glu Asp Asp Leu Leu Gly Asn Ala Lys Pro Arg Tyr Phe Tyr Thr Ser  
180 185 190  
Ala

25  
30  
35  
40  
45  
50  
55  
60

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 205 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 495:

Met Ala Ala Gly Asp Gln Val Phe Ser Gly Ala Gly His Val Xaa Glu  
1 5 10 15  
His Val Ala Gly Gly Arg His Ala Trp Leu Leu Thr Trp Gln Ser Ala  
20 25 30  
Cys Pro Ala Asn Arg Leu Ser Leu Val Pro Leu Val Pro Ser Ala Ser  
35 40 45  
Met Thr Arg Leu Met Arg Xaa Arg Thr Ala Ser Gly Ser Ser Val Ile  
50 55 60  
Leu Trp Met Ala Pro Ala Ala Pro Thr Pro Ala Arg Ala Pro Glu  
65 70 75 80  
Ala Ala Pro Thr Pro Ala Arg Ala Pro Ala Ala Arg Thr Pro Ala  
85 90 95  
Arg Gly Pro Thr Trp Thr Ser Pro Pro Thr Arg Val Leu Leu Gly Thr  
100 105 110

656

Xaa Pro Gly Pro Ser Pro Trp Arg Ser Pro Ala Arg Arg Pro Ala Gln  
115 120 125  
Leu Pro Pro Asp Ser Asp Leu Cys Ser Gly Pro Leu Leu Pro Gly  
130 135 140  
Pro Phe Ser Pro Pro Ala Cys His Thr Ala Pro Asn Ser Val Leu Ile  
145 150 155 160  
Gln Ser Leu Phe Cys Lys Ser Glu Leu Trp Trp Arg Gln Met Arg Ser  
165 170 175  
Ile Thr Trp Val Pro Ser Pro Lys Ala Gly Trp Arg Trp Thr Lys Gly  
180 185 190  
Arg Lys Gln Ala Ser Pro His Arg Ile Leu Phe His Xaa  
195 200 205

20  
25  
30  
35  
40  
45  
50  
55  
60

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 147 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 496:

Met Ala Leu Thr Leu Leu Pro Ser Val Ser Arg Leu Pro Gly Glu Arg  
1 5 10 15  
Met Ala Ala Ser Gly Leu Pro Tyr Val Leu His His Lys Ser Ser Leu  
20 25 30  
Met Lys Val Ile Phe Phe Pro Tyr Pro Val Leu Pro Leu Pro Ala Pro  
35 40 45  
Asn Gly Thr Trp Val Pro Arg Leu Val Leu Gly Leu Gly Ser Gly Asp  
50 55 60  
Gln Val His Tyr Leu Pro Ile Ser Ser Ser Ile Val Asn Tyr Gly Thr  
65 70 75 80  
Ser Val Ser Gly Lys Ser Trp Val Phe Leu Val Tyr Pro Leu His Pro  
85 90 95  
Thr Pro Thr Trp Ser Thr Arg Cys Phe Gln Val Trp Asp Leu Leu Ser  
100 105 110  
Val Glu Leu Pro Asp Lys Gly Glu Gly Asn Thr Arg Arg Ala Ser Gly  
115 120 125  
Val Pro Gly Leu Ser Gln Leu Pro Thr Ser His Lys Pro Ile Lys Gln  
130 135 140  
Glu Tyr Xaa  
145

## (2) INFORMATION FOR SEQ ID NO: 497:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 497:

10 Met Val Trp Val Leu Trp Ser Ala Pro Ser Leu Ala Pro Pro Trp Val 15

Gly Pro Cys Trp Pro Ser Thr Gly Asn Cys Cys Leu Cys Glu Val Gly 30

15 Ala Ala Leu Pro Pro Arg Gly Pro Ser Leu Ser Asp Cys Leu Gly Leu 45

20 Pro Pro Trp Thr Pro Trp Gly Pro Ala Trp Thr Leu Ala Gln Ser Xaa 60

## (2) INFORMATION FOR SEQ ID NO: 498:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 94 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 498:

35 Met Ser Thr Gly Ala Leu Asn Thr Ser Pro Pro Ala Ser Asn Arg Leu 15

Glu Ser Thr Leu Asn Glu Tyr Leu Ile Gln Pro Gln Leu His Cys Ser 30

40 Ser Val Gln Arg Leu Thr Leu Lys Trp Gly Cys Ser Ser Leu Gln Arg 45

45 Asp Gly Gln Ala Val Pro Trp Gly Leu Trp Gln Arg Ala Tyr Pro Ser 60

Leu Pro Thr Leu Pro Ser Asp Leu Leu Arg Pro His Ala Val Thr 80

50 Pro Ser Val Ser Val Ser Val His Thr Cys Glu Ser Ser Xaa 90

## (2) INFORMATION FOR SEQ ID NO: 499:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 499:

Met Phe Leu Ile Phe Val Tyr Phe Leu Lys Xaa Leu Phe Ser Ser Ser 15

5 Leu Pro Phe Leu Trp Leu 20

## (2) INFORMATION FOR SEQ ID NO: 500:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 500:

20 Arg Gly Gly Leu Cys Pro Leu Leu Val Pro Gly Pro Leu Ala Arg Gln 15

Glu Pro Ser Pro Ser Leu Gln Gly Cys Ser Glu Ser Pro Val Gly Met 30

25 Asp

## (2) INFORMATION FOR SEQ ID NO: 501:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 501:

35 Met Gln Phe Leu Leu Thr Ala Phe Leu Leu Val Pro Leu Leu Ala Leu 15

40 Cys Asp Val Pro Ile Ser Leu Gly Phe Ser Pro Ser 25

## (2) INFORMATION FOR SEQ ID NO: 502:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 502:

55 Pro Gly Lys Pro Gln Ala Cys Pro Glu Thr Ser Val Leu Pro 15

## (2) INFORMATION FOR SEQ ID NO: 503:

## (i) SEQUENCE CHARACTERISTICS:



661

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 508:

5 Met Trp Arg Cys Arg Gly Lys Leu Ser Phe Pro Leu Phe Ala Val Val 15  
1 5 10  
Ile Val Ser Cys Arg Lys Asp Gly Pro Asp Ala Ala Ala Pro Ala 30  
20 25  
Val Xaa Lys Lys 35

(2) INFORMATION FOR SEQ ID NO: 509:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 509:

15 Met Ala Leu Val Ala Leu Phe Thr Gln Leu Met Arg Xaa Leu Gly Arg 15  
1 5 10  
Cys Pro Gln

(2) INFORMATION FOR SEQ ID NO: 510:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 510:

35 Met Thr Phe Pro Phe Glu Lys Glu Asn Ser Cys Phe Gln Cys Leu Leu 15  
1 5 10  
Phe Asp Ser Trp Arg Glu Gln Thr Arg Thr Asn Ile Gln Pro Gln Arg 30  
20 25

(2) INFORMATION FOR SEQ ID NO: 511:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 511:

50 Met His Leu Leu Asp Phe Phe Arg Asp Leu Val Leu Leu Val Leu Leu 15  
1 5 10  
Met Thr Phe Pro Phe Glu Lys Glu Asn Ser Cys Phe Gln Cys Leu Leu 15  
1 5 10  
Phe Asp Ser Trp Arg Glu Gln Thr Arg Thr Asn Ile Gln Pro Gln Arg 30  
20 25

(2) INFORMATION FOR SEQ ID NO: 514:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 514:

50 Leu Asn Glu Ser Tyr Val Ser Arg Ala Gly Gly Trp Phe Ser Met Phe 15  
1 5 10  
Xaa Leu Ile Phe Phe Leu Leu Ala Leu Gly Ser Xaa Leu Cys Leu Leu 30  
20 25  
Leu Cys Leu Pro Ser Phe Asn Lys Thr Arg Arg Lys Gln Lys Pro 45  
35 40

662

Ala Leu Leu Asp Ser Phe Trp Leu Glu Val Gln Lys 25  
20  
(2) INFORMATION FOR SEQ ID NO: 512:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 512:

5 Met Cys Leu Ile His Phe Ile Lys Ile Ile Leu Val Phe Ile Leu Lys 15  
1 5 10  
Leu Trp Leu Tyr Ser Gln Lys Cys Pro Lys 25  
20

(2) INFORMATION FOR SEQ ID NO: 513:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 513:

25 Met Ile His Val His Glu Trp Asn Asp Gln Met Leu Met Val Tyr Ile 15  
1 5 10  
Phe Leu Tyr Pro Val Ser Ile Thr Phe Leu Asn Leu Cys Ser Leu Thr 30  
25  
Cys

(2) INFORMATION FOR SEQ ID NO: 514:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 514:

50 Leu Asn Glu Ser Tyr Val Ser Arg Ala Gly Gly Trp Phe Ser Met Phe 15  
1 5 10  
Xaa Leu Ile Phe Phe Leu Leu Ala Leu Gly Ser Xaa Leu Cys Leu Leu 30  
20 25  
Leu Cys Leu Pro Ser Phe Asn Lys Thr Arg Arg Lys Gln Lys Pro 45  
35 40



663

## (2) INFORMATION FOR SEQ ID NO: 515:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 515:

1 Ser Ser Lys Thr Pro Leu Pro Ser Glu Arg Arg Trp Ile Ser Gly Ser 15  
5  
10 Ser Leu Met Ala Pro Arg Pro Trp Leu Leu Gly Ile Ala Leu Leu Gly 30  
20  
15 Leu Trp Ala Leu Glu Pro Ala Leu Gly His Trp 35  
35

## (2) INFORMATION FOR SEQ ID NO: 516:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 516:

1 Leu Asn Trp

## (2) INFORMATION FOR SEQ ID NO: 517:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 517:

1 Phe Ala Phe Cys Ala Glu Leu Met Ile Gln Asn Trp Thr Leu Gly Ala 15  
5  
10 Val Asp Ser Gln Met Asp Asp Met Asp Leu Asp Lys Glu Phe 30  
20  
45 Leu Gln Asp Leu Lys Glu Leu Lys Val Leu Val Ala Asp Lys Asp Leu 45  
35  
50 Leu Asp Leu His Lys Ser Leu Val Cys Thr Ala Leu Arg Gly Lys Leu 60  
55  
65 Gly Val Phe Ser Glu Met Glu Ala Asn Phe Lys Asn Leu Ser Arg Gly 70  
75  
85 Leu Val Asn Val Ala Ala Lys Leu Thr His Asn Lys Asp Val Arg Asp 90  
95  
100 Leu Phe Val Asp Leu Val Glu Lys Phe Val Glu Pro Cys Arg Ser Asp 105  
110

664

His Trp Pro Leu Ser Asp Val Arg Phe Phe Leu Asn Gln Tyr Ser Ala 115  
120  
125

5 Ser Val His Ser Leu Asp Gly Phe Arg His Gln Ala Ser Gly Thr Ala 130  
135  
140

10 Thr Trp Ala Pro Ser Ala Ala Ala Ser Cys Ala Cys Ile Met Thr Glu 145  
150  
155  
160

Val Pro Asn Ala Pro Pro Thr Leu Thr Ile Lys Leu Leu 165  
170

## (2) INFORMATION FOR SEQ ID NO: 518:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 518:

1 Met Trp Lys Asn Leu Gly Ser Gly Ser Val Phe Val Thr Trp Phe Ser 5  
10  
15  
20 Leu Val Met Ile Leu Ser Gly Ile Gly Pro Leu Gly Asp Ala Glu Asp 25  
30  
35 Ser Ile Ser Asp Val Ser His Arg Leu Arg Pro 40

## (2) INFORMATION FOR SEQ ID NO: 519:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 519:

1 Phe Gln Phe Pro Leu Leu Thr Ile Ala Leu Gln Phe Leu 5  
10

## (2) INFORMATION FOR SEQ ID NO: 520:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 520:

1 Met His Tyr Val Ile Val Leu Ser Leu Phe Val Val Leu Glu Lys Lys 5  
10  
15  
20 Asn Lys Met Gly Ser Asp Gly Cys Leu Arg Lys Asn Gly Ser 25  
30

665

666

5

(2) INFORMATION FOR SEQ ID NO: 521:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 521:

Met Ser Arg Ser Ile Val Leu Arg Gly Ser Leu Phe Leu Phe Ser  
1 5 10 15

15

His Tyr Thr Leu Lys Leu Ser Val Ile Lys Gln Thr Asn Arg Lys  
20 25 30

Ile Val Trp Glu Lys Pro Cys Ile Arg Leu Phe Tyr Xaa Val Leu  
35 40 45

20

(2) INFORMATION FOR SEQ ID NO: 522:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 522:

Met Pro Leu Pro Val Leu Leu Cys Leu Thr Leu Pro Met Pro Leu Pro  
1 5 10 15

30

Ser Ala Thr Ala Arg Gly Gly Asn Arg Thr  
20 25

35

(2) INFORMATION FOR SEQ ID NO: 523:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 58 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 523:

Ser Ser Ile Pro Val Ser Ile Leu Ile Gly Met Lys Leu Ile Leu Tyr  
1 5 10 15

50

Leu Leu Ile Thr Glu Ser Gly Ser His Glu Lys Lys Ser Phe Tyr Pro  
20 25 30

Ser Phe Lys Tyr Met Phe Lys Ile Ile Tyr Val Ser Ala Tyr Cys  
35 40 45

55

Arg Thr Ala Leu Arg Ala Thr Val Ser His  
50 55

60

(2) INFORMATION FOR SEQ ID NO: 524:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 524:

Asn Arg Thr Leu Leu Phe Leu Ile Leu Phe Val Leu Phe Gly Leu Gly  
1 5 10 15

Tyr Gly Phe

15

(2) INFORMATION FOR SEQ ID NO: 525:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 525:

Met Phe Leu Leu Val Leu Ser Val Phe Cys Asp Phe Met Cys Ser Ile  
1 5 10 15

Ala Pro Arg Cys His Ala Leu Ser Leu Val Ser Leu Arg Ala Gln His  
20 25 30

30

Leu Ser Leu Phe Ile Thr Cys His  
35 40

35

(2) INFORMATION FOR SEQ ID NO: 526:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 57 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 526:

Met Leu Leu Phe Ile Leu Leu Thr Leu Ser Ser Gly Cys Arg Leu Leu  
1 5 10 15

Val Ser Ser Trp Lys Thr Phe Leu Pro His Phe Ser Leu Pro Gly Pro  
20 25 30

50

Arg Glu His Pro Glu Gly Ser Arg Thr Trp Phe Arg Tyr Trp Glu  
35 40 45

Pro Gly Ala His Cys Leu His Cys Ala  
50 55

60

(2) INFORMATION FOR SEQ ID NO: 527:

- (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 527:  
5 Ala Arg Leu Leu Leu Phe Leu Ser Ser Val His Pro Ser Ile Met Pro  
1 5 10 15  
Ser Cys Asn Gln Leu  
20  
10  
(2) INFORMATION FOR SEQ ID NO: 528:  
15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 528:  
20 Met Ser Leu Thr Ser Ser Leu Thr Phe Leu Ser His Ile Leu Leu Leu  
1 5 10 15  
Pro Gln Lys Leu Gln Phe Leu Ser Trp Met Glu Arg Gln Arg Cys  
25 20 25  
Thr Gly Val Ala Lys Tyr Ala  
35  
30  
(2) INFORMATION FOR SEQ ID NO: 529:  
35 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 128 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 529:  
40 Met Val Leu Arg Leu Ile Gln Leu Ile Phe Leu Ile Phe Phe Ile His  
1 5 10 15  
Ile Ile Ile Leu Leu Ile Pro Gly Ser Arg Pro Cys Gly Ser Trp Val  
45 20 25 30  
Asn Asp Arg Xaa Leu Gly Leu Arg Asp Val Thr His Leu Ile Tyr Leu  
35 40 45  
His Trp Val His Gly His Leu Pro Trp Cys His Pro Tyr Ile Gln Val  
50 55 60  
Glu Phe Ser Ala Leu Ile Glu Ser Thr Ala Gln Leu Gly Leu Pro Phe  
65 70 75 80  
Ser Trp Val Arg Val Ile His Pro Phe Leu Val Leu Pro Cys Leu Tyr  
85 90 95  
Ser Pro Gly Leu Lys Asn(Gly Ile Phe Leu Phe Leu Leu Arg Ala Met  
100 105 110

Pro Gly Gly Met Phe Pro Gly Asn Leu Glu Ala Phe Arg Val Pro Val  
115 120 125  
5  
(2) INFORMATION FOR SEQ ID NO: 530:  
10 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 82 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 530:  
15 Met Gly Ser Ser Val Leu Pro Phe Cys Val Cys Val Thr Ser Pro Ser  
1 5 10 15  
Leu Gly Gly Arg Cys Ile Gln Gly Arg Phe Ala Ser His Ser Lys Phe  
20 25 30  
Trp Gly Phe Gly Xaa Lys Thr Ala Ser Phe Gly Ala Val Gly Glu Thr  
25 35 40 45  
Pro Pro Asp Gln Glu Pro Gln Lys Glu Thr Glu Pro Ala Thr Ser Ser  
50 55 60  
His Ala Arg Pro Trp Ala Arg Val Ile Gly Leu Arg Ile Trp Pro Gln  
65 70 75 80  
Pro Asn  
30  
35  
(2) INFORMATION FOR SEQ ID NO: 531:  
40 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 531:  
45 Met Leu Leu Ser Val Ala Ile Phe Ile Leu Leu Thr Leu Val Tyr Ala  
1 5 10 15  
Tyr Trp Thr Met  
50 20  
55  
(2) INFORMATION FOR SEQ ID NO: 532:  
60 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 75 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 532:

Asn Cys Glu Ile Leu Glu Tyr Cys Tyr Tyr Leu Thr Gln Leu Lys Ile  
1 5 10 15

5 Ser Met Gly Lys Tyr Leu Ser Ile Pro Thr Val Leu Leu Lys Ile Ile  
20 25 30

Arg Cys Ser Ile Thr Ala Val Ser Asp Ser Ser Thr Trp Ala Ile  
35 40 45

10 Lys Ala Gln Leu Lys Ile Glu Asn Lys Asp Leu Asp Asn Lys Thr Ala  
50 55 60

15 Lys Gly Gly Gly Gln Glu Ala Leu Thr Cys Thr  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 533:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 533:

25 Met Phe Leu Met Arg Met His Leu Cys Phe Cys Lys Tyr Cys Cys Ser  
1 5 10 15

30 Phe Ile Val Thr Pro Thr Ser Thr Ser Asn Thr Xaa Ser Tyr Leu Trp  
20 25 30

Pro Trp Ile Ser Ala Ser Met Ala Gly Arg Gly Ser Xaa Trp Ala Cys  
35 40 45

35 Thr Leu Asn Ala Val Thr Arg Glu Gly Leu Pro Glu  
50 55 60

(2) INFORMATION FOR SEQ ID NO: 534:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 534:

45 Met Ser Leu Leu Asn Thr His Thr Leu Cys Phe Val Leu Phe Cys Phe  
1 5 10 15

50 Thr Leu Ser Ile Asn Gln Glu Lys Leu Ala Asn His Leu Ala Phe Arg  
20 25 30

55 Ile Leu Phe Phe Ile Val Phe  
35

(2) INFORMATION FOR SEQ ID NO: 535:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 535:

Met Leu  
1 10

(2) INFORMATION FOR SEQ ID NO: 536:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 536:

20 Met Asp Gln Phe Lys Ile Phe Tyr Phe Leu Lys Ala Phe Phe Ala Cys  
1 5 10 15

25 Cys Asn Val Gln Asp Pro Ser Phe Met Gly Glu Thr Gly Ser Tyr  
20 25 30

Leu Asn Ile Gly  
35

(2) INFORMATION FOR SEQ ID NO: 537:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 537:

40 Met Phe Asp Phe Leu Ser Tyr Phe Lys Asp Leu Leu Ser Cys  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 538:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 538:

Met Gly Phe Gly Phe Val Leu Asn Ile Phe Ser Phe Phe Leu Xaa Pro  
1 5 10 15

Pro Leu

60

671

(2) INFORMATION FOR SEQ ID NO: 539:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 539:

10 Leu Leu Leu Thr Leu Leu Ala Xaa Tyr Xaa  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 540:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 108 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 540:

25 Met Ala Ala Gln Lys Asp Gln Gln Lys Asp Ala Glu Ala Glu Gly Leu  
1 5 10 15

30 Ser Gly Thr Thr Leu Pro Lys Leu Ile Pro Ser Gly Ala Gly Arg  
20 25 30

35 Glu Trp Leu Glu Arg Arg Ala Thr Ile Arg Pro Trp Ser Thr Phe  
35 40 45

40 Val Asp Gln Gln Arg Phe Ser Arg Pro Arg Asn Leu Gly Glu Leu Cys  
50 55 60

35 Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe  
65 70 75 80

Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met Leu  
85 90 95

40 Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Xaa  
100 105

(2) INFORMATION FOR SEQ ID NO: 541:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 106 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 541:

50 Phe Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met  
1 5 10 15

55 Leu Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Tyr Ile Leu Tyr  
20 25 30

60 Leu Arg Thr Leu Glu Ser Lys Leu Val Leu Phe Gly Arg Glu Val Ser  
35 40 45

672

Pro Ala His Gln Tyr Ala Leu Ala Gly Gly Ile Ser Phe Pro Phe Phe  
50 55 60

5 Trp Leu Ala Gly Ala Gly Ser Ala Val Phe Trp Val Leu Gly Ala Thr  
65 70 75 80

Leu Val Val Ile Gly Ser His Ala Ala Phe His Gln Ile Glu Ala Val  
85 90 95

10 Asp Gly Glu Glu Leu Gln Met Glu Pro Val  
100 105

(2) INFORMATION FOR SEQ ID NO: 542:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 136 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 542:

25 Met Asp Arg Phe Thr Val Ala Gly Val Leu Pro Asp Ile Glu Gln Phe  
1 5 10 15

Phe Asn Ile Gly Asp Ser Ser Ser Gly Leu Ile Gln Thr Val Phe Ile  
20 25 30

30 Ser Ser Tyr Met Val Leu Ala Pro Val Phe Gly Tyr Leu Gly Asp Arg  
35 40 45

Tyr Asn Arg Lys Tyr Leu Met Cys Gly Gly Ile Ala Phe Trp Ser Leu  
50 55 60

35 Val Thr Leu Gly Ser Ser Phe Ile Pro Gly Glu His Phe Trp Leu Leu  
65 70 75 80

40 Leu Leu Thr Arg Gly Leu Val Gly Val Gly Glu Ala Ser Tyr Ser Thr  
85 90 95

Ile Ala Pro Thr Leu Ile Ala Asp Leu Phe Val Ala Asp Gln Arg Thr  
100 105 110

45 Gly Cys Ser Ala Ser Ser Thr Leu Pro Phe Arg Trp Ala Val Val Trp  
115 120 125

50 Ala Thr Leu Gln Ala Pro Lys Xaa  
130 135

(2) INFORMATION FOR SEQ ID NO: 543:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 424 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 543:

60

673

674

Met Ala Gly Asp Trp His Trp Ala Leu Arg Val Thr Pro Gly Leu Gly 15  
1 5 10  
Val Val Ala Val Leu Leu Leu Phe Leu Val Val Arg Glu Pro Pro Arg 30  
20 25  
Gly Ala Val Glu Arg His Ser Asp Leu Pro Pro Leu Asn Pro Thr Ser 45  
35 40  
Trp Trp Ala Asp Leu Arg Ala Leu Ala Arg Asn Pro Ser Phe Val Leu 60  
50 55  
Ser Ser Leu Gly Phe Thr Ala Val Ala Phe Val Thr Gly Ser Leu Ala 80  
65 70 75  
Leu Trp Ala Pro Ala Phe Leu Leu Arg Ser Arg Val Val Leu Gly Glu 95  
85 90  
Thr Pro Pro Cys Leu Pro Gly Asp Ser Cys Ser Ser Ser Asp Ser Leu 110  
100 105  
Ile Phe Gly Leu Ile Thr Cys Leu Thr Gly Val Leu Gly Val Gly Leu 125  
115 120  
Gly Val Glu Ile Ser Arg Arg Xaa Arg His Ser Asn Pro Arg Ala Asp 140  
130 135  
Pro Leu Val Cys Ala Thr Gly Leu Leu Gly Ser Ala Pro Phe Leu Phe 160  
145 150 155  
Leu Ser Leu Ala Cys Ala Arg Gly Ser Ile Val Ala Thr Tyr Ile Phe 175  
165 170  
Ile Phe Ile Gly Glu Thr Leu Leu Ser Met Asn Trp Ala Ile Val Ala 190  
180 185  
Asp Ile Leu Leu Tyr Val Val Ile Pro Thr Arg Arg Ser Thr Ala Glu 205  
195 200  
Ala Phe Gln Ile Val Leu Ser His Leu Leu Gly Asp Ala Gly Ser Pro 220  
210 215  
Tyr Leu Ile Gly Leu Ile Ser Asp Arg Leu Arg Arg Asn Trp Pro Pro 240  
225 230 235  
Ser Phe Leu Ser Glu Phe Arg Ala Leu Gln Phe Ser Leu Met Leu Cys 255  
245 250  
Ala Phe Val Gly Ala Leu Gly Gly Ala Leu Ser Trp Ala Pro Xaa Ser 270  
260 265  
Ser Leu Arg Pro Thr Ala Gly Gly His Ser Cys Thr Cys Arg Ala Cys 285  
275 280  
Cys Thr Lys Gln Gly Pro Gln Thr Thr Gly Leu Trp Cys Pro Ser Gly 300  
290 295  
Ala Ala Pro Ala Cys Pro Trp Pro Val Cys Ser Ser Glu Arg Leu 320  
305 310 315

Pro Leu Thr Tyr Leu His Ile Cys His Ser Xaa Pro Trp Ala His Pro 335  
325 330  
Thr Lys Gly Leu Gly Leu Thr Pro Trp Pro Gly Pro Ala Ser Arg Gly 350  
340 345  
Thr Leu Gly Arg Val Pro Ala Pro Arg His Tyr Xaa Gly Ser Ser Gly 365  
355 360  
Glu Glu Val Gly Val Gln Glu Gly Asp Pro Ser Pro Gln Gly Xaa Pro 380  
370 375  
Gln Gly Leu Gly Ala Ile Cys Asn Gly Ile Lys Phe Val Ala Arg Pro 400  
385 390 395  
Gln Val Pro Ala Leu Val Phe Leu Trp Val Ala Ser Asp Leu Ala Pro 415  
405 410  
Arg Leu His Pro Arg Ala Pro Glu 420  
425  
(2) INFORMATION FOR SEQ ID NO: 544:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 544:  
Met Phe Arg Phe Val Ile Cys Leu Phe Leu Trp Leu Val Leu Cys Arg 15  
1 5 10  
Asp Ser Thr Ser Ala Ser Arg Ile Ala Leu Tyr Tyr Arg Ile Val Phe 30  
20 25  
Leu Ile His Gln Cys Ser Ser 35  
35  
(2) INFORMATION FOR SEQ ID NO: 545:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 58 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 545:  
Met Leu Pro Trp Xaa Ala Gln Leu Leu Asp Arg Thr Ile Gly Pro Leu 15  
1 5 10  
Tyr Leu Leu Phe Val Gln Phe Ser Pro Ala Phe Ser Arg Thr Ser Pro 30  
20 25  
Trp Arg Ser Pro Lys Asn Phe Arg Arg Leu Tyr Pro Cys Thr Thr 45  
35 40  
Ser Gly Cys Ala Ala Arg Trp Leu Phe Ser 60

675

50 55

## 5 (2) INFORMATION FOR SEQ ID NO: 546:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 546:

Met Gly Leu Ser Val Leu Leu Pro Leu Cys Leu Leu Gly Pro Gly Arg  
1 5 10 15

Phe Thr Ser Gly Gln Lys Pro Leu Asp Thr Pro Gly Leu Gly Val Pro  
20 25 30

20 Phe

## 25 (2) INFORMATION FOR SEQ ID NO: 547:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 367 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 547:

Met Ala Lys Pro Gln Val Val Val Ala Pro Val Leu Met Ser Lys Leu  
1 5 10 15

Ser Val Asn Ala Pro Gln Phe Tyr Pro Ser Gly Tyr Ser Ser Tyr  
20 25 30

Thr Gln Ser Tyr Gln Asp Gly Cys Gln Asp Tyr Pro Thr Leu Ser Gln  
35 40 45

Tyr Val Gln Asp Phe Leu Asn His Leu Thr Gln Gln Pro Gly Ser Phe  
50 55 60

Gln Thr Gln Ile Gln Gln Phe Ala Gln Thr Leu Asn Gly Cys Val Thr  
65 70 75 80

Thr Asp Asp Ala Leu Gln Gln Leu Val Gln Leu Ile Tyr Gln Gln Ala  
85 90 95

Thr Ser Ile Pro Asn Phe Ser Tyr Met Gly Ala Arg Leu Cys Asn Tyr  
100 105 110

Leu Ser His His Leu Thr Ile Ser Pro Gln Ser Gly Asn Phe Arg Gln  
115 120 125

Leu Leu Leu Gln Arg Cys Arg Thr Gln Tyr Gln Val Lys Asp Gln Ala  
130 135 140

Ala Lys Gly Asp Gln Val Thr Arg Lys Arg Phe His Ala Phe Val Leu  
145 150 155 160

676

Phe Leu Gly Gln Leu Tyr Leu Asn Leu Gln Ile Lys Gly Thr Asn Gly  
165 170 175

Gln Val Thr Arg Ala Asp Ile Leu Gln Val Gly Leu Arg Gln Leu Leu  
180 185 190

Asn Ala Leu Phe Ser Asn Pro Met Asp Asp Asn Leu Ile Cys Ala Val  
195 200 205

Lys Leu Leu Lys Leu Thr Gly Ser Val Leu Gln Asp Ala Trp Lys Gln  
210 215 220

Lys Gly Lys Met Asp Met Gln Gln Ile Ile Gln Arg Ile Gln Asn Val  
225 230 235 240

Val Leu Asp Ala Asn Cys Ser Arg Asp Val Lys Gln Met Leu Leu Lys  
245 250 255

Leu Val Gln Leu Arg Ser Ser Asn Trp Gly Arg Val His Ala Thr Ser  
260 265 270

Thr Tyr Arg Gln Ala Thr Pro Gln Asn Asp Pro Asn Tyr Phe Met Asn  
275 280 285

Gln Pro Thr Phe Tyr Thr Ser Asp Gly Val Pro Phe Thr Ala Ala Asp  
290 295 300

Pro Asp Tyr Gln Gln Lys Tyr Gln Gln Leu Leu Arg Gln Asp Phe  
305 310 315 320

Phe Pro Asp Tyr Gln Gln Asn Gly Thr Asp Leu Ser Gly Ala Gly Asp  
325 330 335

Pro Tyr Leu Asp Asp Ile Asp Asp Gln Met Asp Pro Gln Ile Gln Gln  
340 345 350

Ala Tyr Gln Lys Phe Cys Leu Gln Ser Gln Arg Lys Arg Lys Gln  
355 360 365

## (2) INFORMATION FOR SEQ ID NO: 548:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 548:

Met Leu Arg Leu Asp Ile Ile Asn Ser Leu Val Thr Thr Val Phe Met  
1 5 10 15

Leu Ile Val Ser Val Leu Ala Leu Ile Pro Gln Thr Thr Leu Thr  
20 25 30

Val Gly Gly Gly Val Phe Ala Leu Val Thr Ala Val Cys Cys Leu Ala  
35 40 45

Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro Ser Gly Pro  
60

677

678

50 55 60  
Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu  
65 70 75

5  
Met Leu Lys Gln Val Met Phe Val Phe Ser Gly Met Gly Pro Arg Ser  
1 5 10 15  
His Cys Trp Gly Leu Pro Leu Ala Cys Gly Thr Phe Val Gln Gly His  
20 25 30  
Gln Ala Asp Ser Ser His Leu Leu Pro Leu Lys His Gln Gly Ala  
35 40 45

25  
(2) INFORMATION FOR SEQ ID NO: 550:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 168 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 550:

35 Met Leu Leu Ser Leu Ala Ala Phe Ser Val Ile Ser Val Val Ser Tyr  
1 5 10 15  
Leu Ile Leu Ala Leu Leu Ser Val Thr Ile Ser Phe Arg Ile Tyr Lys  
20 25 30  
Ser Val Ile Gln Ala Val Gln Lys Ser Glu Gly His Pro Phe Lys  
35 40 45

40  
Ala Tyr Leu Asp Val Asp Ile Thr Leu Ser Ser Glu Ala Phe His Asn  
50 55 60  
Tyr Met Asn Ala Ala Met Val His Ile Asn Arg Ala Leu Lys Leu Ile  
65 70 75

50 Ile Arg Leu Phe Leu Val Glu Asp Leu Val Asp Ser Leu Lys Leu Ala  
85 90 95  
Val Phe Met Trp Leu Met Thr Tyr Val Gly Ala Val Phe Asn Gly Ile  
100 105 110  
Thr Leu Leu Ile Leu Ala Glu Leu Leu Ile Phe Ser Val Pro Ile Val  
115 120 125

55 Tyr Glu Lys Tyr Lys Thr Gln Ile Asp His Tyr Val Gly Ile Ala Arg  
130 135 140 145

Asp Gln Thr Lys Ser Ile Val Glu Lys Ile Gln Ala Lys Leu Pro Gly  
145 150 155 160  
5 Ile Ala Lys Lys Lys Ala Glu Xaa  
165

10 (2) INFORMATION FOR SEQ ID NO: 551:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 124 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 551:

15 Ser Val Pro Phe His Leu Leu Val Val Leu Arg Ser Arg Ala Val Arg  
1 5 10 15  
Ala Arg Arg Arg Arg Glu Pro Arg Ser Leu Pro Arg Pro Gly Asp Glu  
20 25 30  
Glu Leu Gln Leu Leu Cys Gly Ala Arg Ser Asp Phe Leu Glu Arg  
35 40 45  
Cys Glu Glu Asp Trp Val Cys Leu Trp His His Ala Asp His Ala Ala  
50 55 60

30 Phe Pro Gly Ser Phe Gln Cys His Gln Cys Gly Phe Leu Pro His Pro  
65 70 75  
Gly Ser Ser Leu Cys His His Gln Leu Gln Asp Leu Val Arg His  
85 90 95  
Pro Ser Cys Thr Glu Val Arg Arg Arg Pro Ser Ile Gln Ser Leu Pro  
100 105 110  
Gly Arg Arg His Tyr Ser Val Leu Arg Ser Phe Pro  
115 120

40  
(2) INFORMATION FOR SEQ ID NO: 552:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 177 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 552:

45 Met Val His Leu Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp  
1 5 10 15  
Val Thr Phe Val Ser Gly Phe Leu Leu Phe Arg Ser Leu Pro Arg His  
20 25 30  
Thr Phe Gly Leu Val Gln Ser Lys Leu Phe Pro Phe Tyr Phe His Ile  
35 40 45



Ser Met Gly Cys Ala Phe Ile Asn Leu Cys Ile Leu Ala Ser Gln His  
50 55 60  
Ala Trp Ala Gln Leu Thr Phe Trp Glu Ala Ser Gln Leu Tyr Leu Leu  
5 65 70 75 80  
Phe Leu Ser Leu Thr Leu Ala Thr Val Asn Ala Arg Trp Leu Glu Pro  
85 90 95  
10 Arg Thr Thr Ala Met Trp Ala Leu Gln Thr Val Glu Lys Glu Arg  
100 105 110  
Gly Leu Gly Gly Glu Val Pro Gly Ser His Gln Gly Pro Asp Pro Tyr  
115 120 125  
15 Arg Gln Leu Arg Glu Lys Asp Pro Lys Tyr Ser Ala Leu Arg Gln Asn  
130 135 140  
Phe Phe Arg Tyr His Gly Leu Ser Ser Leu Cys Asn Leu Gly Cys Val  
145 150 155 160  
Leu Ser Asn Gly Leu Cys Leu Ala Gly Leu Ala Leu Glu Ile Arg Ser  
165 170  
25 Leu  
30 (2) INFORMATION FOR SEQ ID NO: 553:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 72 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 553:  
Met Ala Phe Ile Leu Leu Phe Tyr Cys Leu Met Thr Phe Leu Ser Leu  
1 5 10 15  
Glu Gln Asn Ser Ala Thr Val Glu Pro Ser Ser His Glu Ile Leu His  
20 25 30  
Leu Leu Gln Asn Cys Phe Glu Leu Leu Arg Thr Ser Thr Ser Gln Cys  
35 40 45  
Thr Glu Gly Ile Pro Cys Ala Lys Ile Pro Glu Trp Val Thr His Leu  
50 55 60  
50 Thr Trp Gln Thr Leu Lys Asn Ser  
65 70  
55 (2) INFORMATION FOR SEQ ID NO: 554:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 45 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 554:  
Val Leu Arg Ile Ile Cys Leu Trp Pro Cys Gly Thr Thr Leu Pro Leu  
1 5 10 15  
Val Glu Lys Ala His Asp Ser His Ser Ala Asp Pro Val Cys Pro Gly  
20 25 30  
10 Leu Thr Ala His Leu Pro Val Leu Leu Tyr Val Gln Leu  
35 40 45  
15 (2) INFORMATION FOR SEQ ID NO: 555:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 251 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 555:  
Met Lys His Ala Asp Pro Arg Ile Gln Gly Tyr Pro Leu Met Gly Ser  
1 5 10 15  
25 Pro Leu Leu Met Thr Ser Ile Leu Leu Thr Tyr Val Tyr Phe Val Leu  
20 25 30  
Ser Leu Gly Pro Arg Ile Met Ala Asn Arg Lys Pro Phe Gln Leu Arg  
35 40 45  
Gly Phe Met Ile Val Tyr Asn Phe Ser Leu Val Ala Leu Ser Leu Tyr  
50 55 60  
30 Ile Val Tyr Glu Phe Leu Met Ser Gly Trp Leu Ser Thr Tyr Thr Trp  
65 70 75 80  
Arg Cys Asp Pro Gln Asp Cys Thr Leu Gly Gln Cys Pro Ser Val Pro  
85 90 95  
40 Ser Pro Xaa Thr Pro Val Thr Lys Ala Tyr Val Val Arg Thr Glu Gln  
100 105 110  
Gly Thr Gly Pro Pro Leu Pro Thr Ala Ala Leu Gln Gly Pro Arg Leu  
115 120 125  
45 Trp Phe Leu Thr His Phe Pro Arg Ala Ala Pro Gly Met Trp Pro His  
130 135 140  
50 Cys Cys Leu Pro Leu Gln Ser Trp Gly Leu Lys Gly Leu Tyr Ser Tyr  
145 150 155 160  
Phe Pro Leu Pro Ala Leu Lys Leu Gly Arg Gly Ala Leu Arg Ala Gly  
165 170 175  
55 Pro Thr Lys Gly Leu Val Ala Phe Phe Leu Thr Gln Lys Arg Ser Ala  
180 185 190  
Ile Met Ser Leu Trp Thr Gln Ser His Ser Thr Pro His Thr Glu  
195 200 205  
60

681

Ala Val Ala Ser Gly Pro Lys Val Arg Val Gly Gly Gly Leu Gly Ile  
210 215 220

5 Gln Pro Val Glu Ala Ala Tyr Ser Thr Cys Val Leu Ile Lys Ser Asp  
225 230 235 240

Arg Gly Asn His Glu Lys Lys Lys Lys Lys  
245 250

10

(2) INFORMATION FOR SEQ ID NO: 556:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 556:

20 Gly Leu Ala Gly Leu Cys Gly Gln Leu Ser Ser Pro Ala Leu Cys Val  
1 5 10 15

Asn Arg Leu

25

(2) INFORMATION FOR SEQ ID NO: 557:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 217 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 557:

35 Met Ile Thr Glu Lys Trp Gly Leu Asn Met Glu Tyr Cys Arg Gly Gln  
1 5 10 15

40 Ala Tyr Ile Xaa Ser Ser Gly Phe Ser Ser Lys Met Lys Val Val Ala  
20 25 30

Ser Arg Leu Leu Glu Lys Tyr Pro Gln Ala Ile Tyr Thr Leu Cys Ser  
35 40 45

45 Ser Cys Ala Leu Asn Met Trp Leu Ala Lys Ser Val Pro Val Met Gly  
50 55 60

Val Ser Val Ala Leu Gly Thr Ile Glu Glu Val Cys Ser Phe His  
65 70 75 80

50 Arg Ser Pro Gln Leu Leu Leu Glu Leu Asp Asn Val Ile Ser Val Leu  
85 90 95

55 Phe Gln Asn Ser Lys Glu Arg Gly Lys Glu Leu Lys Glu Ile Cys His  
100 105 110

Ser Gln Trp Thr Gly Arg His Asp Ala Phe Glu Ile Leu Val Glu Leu  
115 120 125

60 Leu Gln Ala Leu Val Leu Cys Leu Asp Gly Ile Asn Ser Asp Thr Asn

682

130 135 140

Ile Arg Trp Asn Asn Tyr Ile Ala Gly Arg Ala Phe Val Leu Cys Ser  
145 150 155 160

5 Ala Val Ser Asp Phe Asp Phe Ile Val Thr Ile Val Val Leu Lys Asn  
165 170 175

10 Val Leu Ser Phe Thr Arg Ala Phe Gly Lys Asn Leu Gln Gly Gln Thr  
180 185 190

Ser Asp Val Phe Phe Ala Ala Gly Ser Leu Thr Ala Val Leu His Ser  
195 200 205

15 Leu Asn Glu Val Ile Gly Lys Tyr Xaa  
210 215

20 (2) INFORMATION FOR SEQ ID NO: 558:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 82 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 558:

Leu Leu Lys Val Leu Cys Ile Leu Pro Val Met Lys Val Glu Asn Glu  
1 5 10 15

30 Arg Tyr Glu Asn Gly Arg Lys Arg Leu Lys Ala Tyr Leu Arg Asn Thr  
20 25 30

35 Leu Thr Asp Gln Arg Ser Ser Asn Leu Ala Leu Leu Asn Ile Asn Phe  
35 40 45

Asp Ile Lys His Asp Leu Asp Leu Met Val Asp Thr Tyr Ile Lys Leu  
50 55 60

40 Tyr Thr Ser Lys Ser Glu Leu Pro Thr Asp Asn Ser Glu Thr Val Glu  
65 70 75 80

Asn Thr

45

(2) INFORMATION FOR SEQ ID NO: 559:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 559:

Met Val Leu Ile Leu Leu Asn Leu Leu Leu Gly Gln Phe Ser Cys Met  
1 5 10 15

Ser Pro Ala Ser His His Cys His Pro Leu Pro Thr Glu Met Pro Cys  
20 25 30

Ser Ser Asp Trp Gly Phe Asp Ser His Thr Val Tyr Pro Ser Cys Val  
35 40 45  
Asp Ala Leu Leu Pro Lys Pro Ser Ala Asn Ser Phe Pro Asn Gly Ser  
50 55 60  
Cys His Cys Gln Gly Leu Tyr Asn Gln Gln Gln Asn Leu His Ala  
65 70 75 80  
Ala Glu Gly Pro Ala Ser Leu Arg Cys Asn Lys Tyr Val Ser Thr  
85 90 95

(2) INFORMATION FOR SEQ ID NO: 560:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 54 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 560:

Met Ile Pro Ala Tyr Ser Lys Asn Arg Ala Tyr Ala Ile Phe Ile  
1 5 10 15  
Val Phe Thr Val Ile Gly Asp Ala Pro Gly Ala Val Leu Ser Cys Ala  
20 25 30  
Gly His Pro Cys Val Gly Phe Ala Ala Val Leu Val Ala Pro Leu Thr  
35 40 45  
Val Ala Val Ser Ser Xaa  
50

(2) INFORMATION FOR SEQ ID NO: 561:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 108 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 561:

Met Glu Val Pro Pro Ala Pro Arg Ser Phe Leu Cys Arg Ala Leu  
1 5 10 15  
Cys Leu Phe Pro Arg Val Phe Ala Ala Glu Ala Val Thr Ala Asp Ser  
20 25 30  
Glu Val Leu Glu Glu Arg Gln Lys Arg Leu Pro Tyr Val Pro Glu Pro  
35 40 45  
Tyr Tyr Pro Glu Ser Gly Trp Asp Arg Arg Leu Glu Leu Phe Gly Lys  
50 55 60  
Asp Thr Val Asn Thr Ser Leu Asn Val Tyr Arg Asn Lys Asp Ala Leu  
65 70 75 80

Ser His Phe Val Ile Ala Gly Ala Val Thr Gly Ser Leu Phe Arg Ile  
85 90 95  
Asn Val Gly Leu Arg Gly Trp Trp Leu Val Ala Xaa  
100 105

(2) INFORMATION FOR SEQ ID NO: 562:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 50 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 562:

Met Asn Trp Gly Leu Ser Ile Trp Leu His Tyr Tyr Glu Lys Lys Lys  
1 5 10 15  
Glu Gln Val Phe Leu Val Ile Leu Ala His Val Val Arg Arg Cys Ala  
20 25 30  
Ser Asp Gly Ile Leu Gln Phe Glu Ser Ser Leu Leu Lys Met Arg Arg  
35 40 45  
Ala Pro  
50

(2) INFORMATION FOR SEQ ID NO: 563:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 253 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 563:

Met Val Lys Val Cys Asn Asp Ser Asp Arg Trp Ser Leu Ile Ser Leu  
1 5 10 15  
Ser Asn Asn Ser Gly Lys Asn Val Glu Leu Lys Phe Val Asp Ser Leu  
20 25 30  
Arg Arg Gln Phe Glu Phe Ser Val Asp Ser Phe Gln Ile Lys Leu Asp  
35 40 45  
Ser Leu Leu Leu Phe Tyr Glu Cys Ser Glu Asn Pro Met Thr Glu Thr  
50 55 60  
Phe His Pro Thr Ile Ile Gly Glu Ser Val Tyr Gly Asp Phe Gln Glu  
65 70 75 80  
Ala Phe Asp His Leu Cys Asn Lys Ile Ile Ala Thr Arg Asn Pro Glu  
85 90 95  
Glu Ile Arg Gly Gly Leu Leu Lys Tyr Cys Asn Leu Leu Val Arg  
100 105 110  
Gly Phe Arg Pro Ala Ser Asp Glu Ile Lys Thr Leu Gln Arg Tyr Met  
60

685

686

115 120 125

Cys Ser Arg Phe Ile Asp Phe Ser Asp Ile Gly Glu Gln Arg  
130 135 140

5 Lys Leu Glu Ser Tyr Leu Gln Asn His Phe Val Gly Leu Glu Asp Arg  
145 150 155

10 Lys Tyr Glu Tyr Leu Met Thr Leu His Gly Val Asn Glu Ser Thr  
165 170 175

Val Cys Leu Met Gly His Glu Arg Gln Thr Leu Asn Leu Ile Thr  
180 185 190

15 Met Leu Ala Ile Arg Val Leu Ala Asp Gln Asn Val Ile Pro Asn Val  
195 200 205

Ala Asn Val Thr Cys Tyr Tyr Gln Pro Ala Pro Tyr Val Ala Asp Ala  
210 215 220

20 Asn Phe Ser Asn Tyr Tyr Ile Ala Gln Val Gln Pro Val Phe Thr Cys  
225 230 235 240

25 Gln Gln Gln Thr Tyr Ser Thr Trp Leu Pro Cys Asn Xaa  
245 250

(2) INFORMATION FOR SEQ ID NO: 564:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 564:

Met Ser Phe Leu Met Trp Leu Met Ser Leu Ala Ile Thr Ser Gln Pro  
1 5 10 15

40 Pro Met

(2) INFORMATION FOR SEQ ID NO: 565:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 565:

Met Ala Pro Lys Gly Lys Val Gly Thr Arg Gly Lys Lys Gln Ile Phe  
1 5 10 15

55 Glu Glu Asn Arg Glu Thr Thr Leu Lys Phe Tyr Leu Arg Ile Ile Leu Gly  
20 25 30

60 Ala Asn Ala Ile Tyr Cys Leu Val Thr Leu Val Phe Phe Tyr Ser Ser  
35 40 45

Ala Ser Phe Trp Ala Trp Leu Ala Leu Gly Phe Ser Leu Ala Val Tyr  
50 55 60

5 Gly Ala Ser Tyr His Ser Met Ser Met Ala Arg Ala Ala Phe Phe  
65 70 75 80

10

(2) INFORMATION FOR SEQ ID NO: 566:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 73 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 566:

20 His Leu Lys Asp Val Ile Leu Leu Thr Ala Ile Val Gln Val Leu Ser  
1 5 10 15

25 Cys Phe Ser Leu Tyr Val Trp Ser Phe Trp Leu Leu Ala Pro Gly Arg  
20 25 30

Ala Leu Tyr Leu Leu Trp Val Asn Val Leu Gly Pro Trp Phe Thr Ala  
35 40 45

30 Asp Ser Gly Thr Pro Ala Pro Glu His Asn Glu Lys Arg Gln Arg Arg  
50 55 60

Gln Glu Arg Arg Gln Met Lys Arg Leu  
65 70

(2) INFORMATION FOR SEQ ID NO: 567:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 263 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 567:

45 Met Asp Cys Pro Ala Leu Pro Gly Trp Lys Lys Glu Glu Val Ile  
1 5 10 15

50 Arg Lys Ser Gly Leu Ser Ala Gly Lys Ser Asp Val Tyr Tyr Phe Ser  
20 25 30

Pro Ser Gly Lys Lys Phe Arg Ser Lys Pro Gln Leu Ala Arg Tyr Leu  
35 40 45

55 Gly Asn Thr Val Asp Leu Ser Ser Phe Asp Phe Arg Thr Gly Lys Met  
50 55 60

Met Pro Ser Lys Leu Gln Lys Asn Lys Gln Arg Leu Arg Asn Asp Pro  
65 70 75 80

Leu Asn Gln Asn Lys Gly Lys Pro Asp Leu Asn Thr Thr Leu Pro Ile  
85 90 95  
Arg Gln Thr Ala Ser Ile Phe Lys Gln Pro Val Thr Lys Val Thr Asn  
100 105 110  
His Pro Ser Asn Lys Val Lys Ser Asp Pro Gln Arg Met Asn Gln Gln  
115 120 125  
Pro Arg Gln Leu Phe Trp Gln Lys Arg Leu Gln Gly Leu Ser Ala Ser  
130 135 140  
Asp Val Thr Gln Gln Ile Ile Lys Thr Met Gln Leu Pro Lys Gly Leu  
145 150 155 160  
Gln Gly Val Gly Pro Gly Ser Asn Asp Gln Thr Leu Leu Ser Ala Val  
165 170 175  
Ala Ser Ala Leu His Thr Ser Ser Ala Pro Ile Thr Gly Gln Val Ser  
180 185 190  
Ala Ala Val Gln Lys Asn Pro Ala Val Trp Leu Asn Thr Ser Gln Pro  
195 200 205  
Leu Cys Lys Ala Phe Ile Val Thr Asp Gln Asp Ile Arg Lys Gln Gln  
210 215 220  
Glu Arg Val Gln Gln Val Arg Lys Lys Leu Gln Gln Ala Leu Met Ala  
225 230 235 240  
Asp Ile Leu Ser Arg Ala Ala Asp Thr Gln Gln Met Asp Ile Gln Met  
245 250 255  
Asp Ser Gly Asp Gln Ala Xaa  
260  
(2) INFORMATION FOR SEQ ID NO: 568:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 70 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 568:  
Met Met Arg Pro Phe Tyr Leu Leu Leu Pro Val Leu Cys Thr Gln Ala  
1 5 10 15  
Leu Arg Gln Ser Gln Gly Lys Ser Pro Leu Leu Trp Lys Arg Thr Leu  
20 25 30  
Leu Phe Gly Leu Thr His Leu Asn Pro Ser Ala Lys Leu Leu Leu Ser  
35 40 45  
Gln Met Lys Thr Ser Gly Asn Arg Lys Ser Gln Tyr Ser Lys Tyr Ala  
50 55 60  
Arg Asn Trp Lys Lys His  
65 70

(2) INFORMATION FOR SEQ ID NO: 569:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 569:  
Met Pro Val Thr Ser Lys Arg Thr Leu Phe Phe Pro Asp Pro Cys Ser  
1 5 10  
Tyr Asp Thr Pro Pro Asp Cys His Cys His Ser Phe Arg Ala Gln  
20 25 30  
Leu Leu  
20  
(2) INFORMATION FOR SEQ ID NO: 570:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 104 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 570:  
Met Asn Ser Arg Gly Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu  
1 5 10 15  
Leu His Ile Val Leu Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala  
20 25 30  
Trp Thr Leu Thr Asn Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu  
35 40 45  
His Ala Val Lys Gly Thr Pro Phe Gln Thr Pro Asp Gln Gly Lys Ser  
50 55 60  
Lys Ala Pro Asn Ser Leu Gly Thr Thr Gly Leu Trp Ser Thr Val Tyr  
65 70 75 80  
Ile Phe Thr Gln Val Phe His Asn Phe Ser Asn Asn Ser Ile Phe Ser  
85 90 95  
Gly Lys Phe Leu Tyr Gln Val Xaa  
100

(2) INFORMATION FOR SEQ ID NO: 571:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 132 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 571:

689

690

Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu Leu His Ile Val Leu  
1 10 15

5 Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala Trp Thr Leu Thr Asn  
20 25 30

10 Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu His Ala Val Lys Gly  
35 40 45

Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala Arg Leu Leu Thr His  
50 55 60

15 Trp Glu Gln Leu Asp Tyr Gly Val Gln Phe Thr Ser Arg Lys Phe  
65 70 75 80

Phe Thr Ile Ser Pro Ile Ile Leu Tyr Phe Leu Ala Ser Phe Tyr Thr  
85 90 95

20 Lys Tyr Asp Pro Thr His Phe Ile Leu Asn Thr Ala Ser Leu Leu Ser  
100 105 110

25 Val Leu Ile Pro Lys Met Pro Gln Leu His Gly Val Arg Ile Phe Gly  
115 120 125

Ile Asn Lys Tyr  
130

(2) INFORMATION FOR SEQ ID NO: 572:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 572:

40 Met Asn Lys Trp Ile Cys Glu Met His Cys Tyr Leu Val Leu Leu Ser  
5 10 15

Val Cys Ser Pro Ser Ala Leu Arg Arg Val Arg His Thr Leu Ser Arg  
20 25 30

45

(2) INFORMATION FOR SEQ ID NO: 573:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 573:

55 Met Pro Val Leu Ser Leu Leu Cys Thr Leu Ile Val Ser Phe Gln Ser  
1 5 10 15

60

Ala Asp Ser Cys Glu Val Phe Leu Asn Cys Ser Leu  
20 25

5

(2) INFORMATION FOR SEQ ID NO: 574:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 574:

15 Met Lys Val Ser Thr Met Leu Trp Phe Leu Cys Trp Glu Gln Ser His  
1 5 10 15

Phe Leu Arg Glu Trp Glu Asp Leu Ser Thr Phe Leu Ile Leu Ile Gln  
20 25 30

20 Met Glu Cys Gln Tyr Gly Asn Ser  
35 40

(2) INFORMATION FOR SEQ ID NO: 575:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 575:

35 Met Gly Leu Pro Leu Met Ala Leu Met Trp Ser Thr Leu Pro Ala Ser  
1 5 10 15

Ala Gly Val Asn Phe Ile Leu Ala Leu Pro Leu Leu Xaa Leu  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 576:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 576:

50 Met Lys Arg Gly Cys Leu Gly Leu Phe Phe Ser Cys Ser Ser  
1 5 10 15

Ala Pro Thr Met Leu Leu Cys Asp Tyr Leu Asn Trp Phe  
20 25

(2) INFORMATION FOR SEQ ID NO: 577:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids

691

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 577:

5 Met Lys Leu Leu Gly Ile Ala Leu Ala Tyr Val Ala Ser Val  
1 5 10  
Tyr Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu Asn Gly Glu Leu  
20 25 30  
Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu Pro Leu Arg Glu Lys  
35 40 45  
Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val Lys  
50 55 60  
Phe Leu Ser Glu Lys Asp Arg Lys Arg Ile Leu Xaa Asn Arg Arg Arg  
65 70 75 80  
Xaa Val Arg Gly Leu Pro Ser Xaa Leu Thr Asn Ser  
85 90

25 (2) INFORMATION FOR SEQ ID NO: 578:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 42 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 578:

30 Met Lys Phe Ser Leu Val Leu Leu Ile Lys Ile Ile Ser Phe Glu Arg  
1 5 10 15  
Leu Leu Ile Phe Leu Phe Pro Leu Ser Phe Leu Pro Asn Ile Tyr Arg  
20 25  
Arg Val Met Val Asn Leu Asn Ile Leu Phe  
35 40

45 (2) INFORMATION FOR SEQ ID NO: 579:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 70 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 579:

50 Leu Ala Gln Glu Cys Pro Pro His Ile Pro Ser Ser Phe Phe Leu Val  
1 5 10 15  
Lys Leu Leu Phe Ile Pro Tyr Leu Ala Ser Leu Leu Pro Leu Ser  
20 25  
Thr Phe Thr Ser Asp Phe Tyr Phe Met Glu Phe Gly Ile Glu Val Lys  
35 40 45

60

692

Leu Gln Gln Cys Arg Gln His Gln Val Leu Gln Glu Lys Asn Thr Lys  
50 55 60

5 Lys Phe Asn Lys Lys Lys  
65 70

10 (2) INFORMATION FOR SEQ ID NO: 580:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 110 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 580:

15 Met Leu Arg Leu Leu Leu Val Ala Phe Ala Leu Val Val Val Leu  
1 5 10 15  
Phe His Val Leu Leu Ala Pro Ile Thr Ala Leu Phe His Thr His Phe  
20 25 30  
Tyr Asp Arg Leu Gln Asp Ala Gly Ser Arg Tyr Pro Glu Leu Tyr Leu  
35 40 45  
Tyr Ser Arg Ala Asp Glu Val Val Leu Ala Arg Asp Ile Glu Arg Met  
50 55 60  
Val Glu Ala Arg Leu Ala Arg Arg Val Leu Ala Arg Ser Val Asp Phe  
65 70 75 80  
Val Ser Ser Ala His Val Ser His Leu Arg Asp Tyr Pro Thr Tyr Tyr  
85 90 95  
Thr Ser Leu Cys Val Asp Phe Met Arg Asn Cys Val Arg Cys  
100 105 110

40 (2) INFORMATION FOR SEQ ID NO: 581:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 581:

45 Met Phe Lys Leu Glu Glu Cys Gly Lys Thr Thr Phe Leu Leu Ser Met  
1 5 10 15  
Ala Leu Tyr Phe Thr Tyr Ile Val Gln Thr Thr Lys Gly Cys  
20 25 30

55 (2) INFORMATION FOR SEQ ID NO: 582:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 71 amino acids  
(B) TYPE: amino acid

60

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 582:

Met Glu Ser Asp Ala Leu Leu Thr Ile Phe Trp Ile Ile Ala Arg  
1 5 10 15

Ser Ser Val Arg Ser Val Gly Lys Ser Ser Gln Arg Ser Phe Thr Thr  
20 25 30

Ile Thr Gln Leu Arg Ser Thr His Thr Gly Pro Ser Arg Arg Ser Tyr  
35 40 45

Leu Ile Trp Trp Asn Gly Gly Pro Lys Arg Thr Ile Ser Tyr Val Ser  
50 55 60

Arg Arg Phe Arg Ser Phe Arg  
65 70

(2) INFORMATION FOR SEQ ID NO: 583:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 583:

Val Gly Leu Phe Gln Pro Lys Thr Phe Gln Val Pro Val Thr Asp Leu  
1 5 10 15

Tyr Ile Phe Ile Lys Ile Tyr Ser Glu Ile Gly Pro Ile Met His Val  
20 25 30

Leu Cys Pro Gly Tyr Ser Gln Ser Pro Ser Thr Pro Pro Trp Thr  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 584:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 584:

Met Trp Phe Gly Ser Asp Arg Ser Asp Leu Arg Ile Gly Thr Ala Phe  
1 5 10 15

Leu Phe Asp Leu Val Cys Asp Leu Cys Ile His Ala Trp Lys Pro Pro  
20 25 30

Gly Leu Val Arg Phe Ser Phe  
35

(2) INFORMATION FOR SEQ ID NO: 585:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 77 amino acids  
(B) TYPE: amino acid

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 585:

Met Leu Asn Thr Ala Ser Leu Asn Leu Pro Trp Lys Val Gln Leu Phe  
1 5 10 15

Ala His Ala  
10

(2) INFORMATION FOR SEQ ID NO: 586:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 586:

Met Ser Ala Cys Leu Leu Phe Leu Ala Phe Ser Trp Lys Arg Lys  
1 5 10 15

Gly Leu Trp Ser Gly Pro Gly  
20 25

(2) INFORMATION FOR SEQ ID NO: 587:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 69 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 587:

Met Leu Pro Pro Phe Ser Leu Val Tyr Thr His Phe Leu Val Ala Ser  
1 5 10 15

Leu Leu Pro Val Ile Leu Ala Val Phe Pro Asp Ser Ala Gln Ile Val  
20 25 30

Pro Leu Leu Lys Pro Ile Pro Arg Pro Gln Pro Glu Val Ile Phe Pro  
35 40 45

Ser Ser Glu Leu Leu Glu Gln Leu Leu Ser Val Gln Phe Val Trp Gln  
50 55 60

Ala His Thr Val Ala  
65

(2) INFORMATION FOR SEQ ID NO: 588:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 77 amino acids  
(B) TYPE: amino acid



695

(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 588:

5 Met Gly Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu  
1 5 10 15  
Leu Met Met Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu  
20 25 30  
10 Ser Thr Cys Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu  
35 40 45  
Asn Val Gly Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr  
50 55 60  
15 Arg Lys Lys Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys  
65 70 75

20 (2) INFORMATION FOR SEQ ID NO: 589:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 155 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 589:

30 Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe Phe  
1 5 10 15  
Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val  
20 25 30  
35 Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro  
35 40 45  
Leu Lys Val Phe Gly Tyr Gln Pro Asp Pro Leu Asn Tyr Gln Ile Ala  
50 55 60  
40 Val Gly Phe Leu Glu Leu Leu Ala Gly Leu Leu Leu Val Met Gly Pro  
65 70 75  
45 Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu Leu Met Met  
85 90 95  
Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu Ser Thr Cys  
100 105 110  
50 Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Leu Asn Val Gly  
115 120 125  
Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys  
130 135 140  
55 Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys Xaa  
145 150 155

60

696

(2) INFORMATION FOR SEQ ID NO: 590:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 590:

5 Met Pro Glu Thr Arg Leu Gly His Arg Gln Gln Phe Ala Val Phe His  
1 5 10 15  
10 Leu Xaa Pro Val Pro Pro Cys Gly  
20

15 (2) INFORMATION FOR SEQ ID NO: 591:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 38 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 591:

20 Met Leu Thr Phe Leu Phe Ser Ala Cys Ala Thr Cys Leu Gly Lys Leu  
1 5 10 15  
25 Ala Ser Pro Leu Ala Pro Val Gly Pro Gln Gln Arg Gly Xaa Pro Pro  
20 25 30  
30 Gly Pro Pro Leu Leu Ser  
35

35 (2) INFORMATION FOR SEQ ID NO: 592:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 69 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 592:

40 Met Asp Pro Phe His Tyr Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu  
1 5 10 15  
45 Val Phe Ala Val Val Leu Phe Ser Val Gly Ile Leu Leu Ile Leu Ser  
20 25 30  
50 Arg Arg Cys Lys Cys Ser Phe Asn Gln Lys Pro Arg Ala Pro Gly Asp  
35 40 45  
Glu Glu Ala Gln Val Glu Asn Leu Ile Thr Ala Asn Ala Thr Glu Pro  
50 55 60  
55 Gln Lys Ala Glu Asn  
65

60

697

## (2) INFORMATION FOR SEQ ID NO: 593:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 308 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 593:

5

10 Asn Leu Arg Val Arg Leu Gly Asp Val Ile Ser Ile Gln Pro Cys Pro  
1 5 10 15

Asp Val Lys Tyr Gly Lys Arg Ile His Val Leu Pro Ile Asp Asp Thr  
20 25 30

15 Val Glu Gly Ile Thr Gly Asn Leu Phe Glu Val Tyr Leu Lys Pro Tyr  
35 40 45

Phe Leu Glu Ala Tyr Arg Pro Ile Arg Lys Gly Asp Ile Phe Leu Val  
50 55 60

20 Arg Gly Gly Met Arg Ala Val Glu Phe Lys Val Val Glu Thr Asp Pro  
65 70 75 80

25 Ser Pro Tyr Cys Ile Val Ala Pro Asp Thr Val Ile His Cys Glu Gly  
85 90 95

Glu Pro Ile Lys Arg Glu Asp Glu Glu Ser Leu Asn Glu Val Gly  
100 105 110

30 Tyr Asp Asp Ile Gly Gly Cys Arg Lys Gln Leu Ala Gln Ile Lys Glu  
115 120 125

Met Val Glu Leu Pro Leu Arg His Pro Ala Leu Phe Lys Ala Ile Gly  
130 135 140

35 Val Lys Pro Pro Arg Gly Ile Leu Leu Tyr Gly Pro Pro Gly Thr Gly  
145 150 155 160

Lys Thr Leu Ile Ala Arg Ala Val Ala Asn Glu Thr Gly Ala Phe Phe  
165 170 175

40 Phe Leu Ile Asn Gly Pro Glu Ile Met Ser Lys Leu Ala Gly Glu Ser  
180 185 190

45 Glu Ser Asn Leu Arg Lys Ala Phe Glu Glu Ala Glu Lys Asn Ala Pro  
195 200 205

Ala Ile Ile Phe Ile Asp Glu Leu Asp Ala Ile Ala Pro Lys Arg Glu  
210 215 220

50 Lys Thr His Gly Glu Val Glu Arg Arg Ile Val Ser Gln Leu Leu Thr  
225 230 235 240

Leu Met Asp Gly Leu Lys Gln Arg Ala His Val Ile Val Met Ala Ala  
245 250 255

Thr Asn Arg Pro Asn Ser Ile Asp Pro Ala Leu Arg Arg Phe Gly Arg  
260 265 270

60 Phe Asp Arg Glu Val Asp Ile Gly Ile Pro Asp Ala Thr Gly Arg Leu

698

275 280 285

Glu Ile Leu Gln Ile His Thr Lys Asn Met Lys Leu Ala Asp Asp Val  
290 295 300

5

Asp Leu Glu Gln  
305

10

(2) INFORMATION FOR SEQ ID NO: 594:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 594:

15 Met Gln Ile Lys Leu Lys Ser Val Lys Thr Val Phe Ala Ile Thr  
1 5 10 15

20 Leu Leu Val Leu Phe Leu  
20

25

(2) INFORMATION FOR SEQ ID NO: 595:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 595:

30 Met Phe Pro Lys Phe Cys Pro Ile Leu Ser Leu Val Asp Phe Ile Ser  
1 5 10 15

His Arg Asp Lys Pro Glu Thr Glu  
20

40

(2) INFORMATION FOR SEQ ID NO: 596:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 596:

45 Met Leu Ile Glu Cys Ala Trp Gln Leu Met Phe Leu Leu Lys Val  
1 5 10 15

Glu Gln Leu Gly Ile Leu Asp Lys  
20

55

(2) INFORMATION FOR SEQ ID NO: 597:

60

699

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 597:

Met  
1

10 (2) INFORMATION FOR SEQ ID NO: 598:  
(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 598:

20 Met Cys Ile Met Ser Ala Leu Val  
1 5

25 (2) INFORMATION FOR SEQ ID NO: 599:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 599:

30 Met Phe Leu Val Trp Phe Trp Gly Leu Ile Ser Ala Leu Ser Asn  
1 5 10 15

35 Val His Thr Pro Ser Arg Leu Pro Ala  
20 25

40 (2) INFORMATION FOR SEQ ID NO: 600:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 600:

50 Met Xaa Gly Leu Ser Leu Ile Leu Thr Val Thr Leu Leu Ala Val Ser  
1 5 10 15

Asp Ser Ala Ala Thr Cys Ile Val Ala Lys Gly  
20 25

55 (2) INFORMATION FOR SEQ ID NO: 601:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 61 amino acids

700

- (B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 601:

5 Met Trp Thr Arg Ser Ser Arg Cys Leu Leu Cys Ile Pro Gly Xaa  
1 5 10 15

Ser Arg Arg Arg Arg Ala Gly Ser Gly Met Lys Pro Arg Ser Trp Ser  
20 25 30

Ala Trp Arg Pro Ser Gly Gly Thr Gly Thr Ser Ser Gln Ser Ser  
35 40 45

15 Thr Gln Ser Arg Thr Leu Ser Ala Thr Ala Ser Pro Ala  
50 55 60

20 (2) INFORMATION FOR SEQ ID NO: 602:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 602:

25 Met Arg Glu Thr Ser Ile Arg Val Leu Leu Met Leu Pro Ala Leu Glu  
1 5 10 15

30 Ser Thr Ser Gly Leu Ser Ala Phe Met Gly Leu Gly Thr  
20 25

35 (2) INFORMATION FOR SEQ ID NO: 603:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 69 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 603:

40 Met Pro Pro Lys Gln Glu Leu Gly Ser Gly Val Gly Glu Leu Ala Lys  
1 5 10 15

45 Asn Ser Lys Arg Gln His Trp Asn His Arg Trp Lys Tyr Leu Lys  
20 25 30

50 Leu Ile Arg Trp Glu Asp Gly Leu Leu Leu Glu Gly Leu Leu Val  
35 40 45

Leu Glu His Cys Ala Thr Met Ala Trp Asp Cys Leu Met Arg Leu Glu  
50 55 60

55 Leu Leu Lys Arg Leu  
65

60 (2) INFORMATION FOR SEQ ID NO: 604:

701

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 604:

5 Lys Ile Val Tyr Ile Leu Gly Asn Pro Leu Lys Phe Asn Ser Arg Val  
1 5 10 15

10 Ile His His Leu Val Leu Leu Gln  
20

## (2) INFORMATION FOR SEQ ID NO: 605:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 605:

25 Met Asn Leu His Gln Arg Arg Leu Leu Ile Gly His Leu Met Thr  
1 5 10 15

30 Leu Val Lys Ala Ser Lys Ser Phe Ser Phe Thr Glu Ile Thr Ser Ser  
20 25 30

Arg Lys Lys  
35

## (2) INFORMATION FOR SEQ ID NO: 606:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 130 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 606:

40 Leu Leu Gly Tyr Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr  
1 5 10 15

45 Tyr Asn Ile His Leu His Ala Leu Phe Tyr Leu Phe Thr Leu Leu Val  
20 25 30

50 Gly Gly Leu Ser Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr  
35 40 45

Val Gly Pro Thr Gln Arg Leu Leu Cys Gly Thr Leu Ala Ala Leu  
50 55 60

55 His Met Leu Phe Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Xaa  
65 70 75 80

Glu Gly Ile Leu Asp Thr Leu Glu Gly Pro Asn Ile Pro Pro Ile Gln  
85 90 95

60

702

Arg Val Pro Arg Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro  
100 105 110

5 Thr Thr Val Leu Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln  
115 120 125

Ser His  
130

10

## (2) INFORMATION FOR SEQ ID NO: 607:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 607:

20 Met Leu Val Ile Phe Leu Phe Thr Ser Leu Leu Lys Ile Pro Ser Ser  
1 5 10 15

25 Val Pro Gly Leu Ile Asn Val  
20

## (2) INFORMATION FOR SEQ ID NO: 608:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 608:

35 Glu Leu Asp Tyr Ile Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO: 609:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 232 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 609:

50 Met Ala Pro Pro Gly Trp Gln Xaa Xaa Xaa Xaa Trp Leu Ala Cys  
1 5 10 15

Pro Asp Arg Gly Glu Leu Ser Ser Arg Ser Pro Pro Cys Arg Leu Ala  
20 25 30

55 Arg Trp Ala Glu Gly Asp Arg Glu Thr Arg Thr Cys Leu Leu Glu Leu  
35 40 45

Ser Ala Gln Ser Trp Gly Gly Arg Phe Arg Arg Ser Ser Ala Val Ser  
50 55 60

60

703

- Ala Gly Ser Pro Ser Arg Leu His Phe Leu Pro Gln Pro Leu Leu Leu  
65 70 75 80  
Arg Ser Ser Gly Ile Pro Ala Ala Ala Thr Pro Trp Pro Gln Pro Ala  
85 90 95  
Gly Leu Pro Val Arg Pro Thr Pro Thr Arg Thr Gly Gln Asp Arg  
100 105 110  
10 Thr Leu Asp Ile Ser Ile Cys Thr Gln Val Leu Ala Gly Thr Gln Gln  
115 120 125  
Pro Pro Pro Arg Met Thr Ser Pro Ser Ser Ser Pro Val Phe Arg  
130 135 140  
15 Leu Gln Thr Leu Asp Gly Gly Gln Gln Asp Gly Ser Gln Ala Asp Arg  
145 150 155 160  
Gly Lys Leu Asp Phe Gly Ser Gly Leu Pro Pro Met Gln Ser Gln Phe  
165 170 175  
20 Gln Gly Gln Asp Arg Lys Phe Ala Pro Ser Asp Lys Ser Gln Pro Pro  
180 185 190  
25 Thr Thr Gln Arg Gln Gln Val Pro Val Ser Arg Ile Gln Thr Asp Leu  
195 200 205  
Thr Gln Ile Gly Ser Ser Met Arg Ser Pro Gly Val Ser Pro Arg Ile  
210 215 220  
30 Trp Leu Asp Phe Gln Ser Thr Xaa  
225 230  
35 (2) INFORMATION FOR SEQ ID NO: 610:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 610:  
Met Val Leu Leu Leu Leu Ala Tyr Val Leu Leu Thr Tyr Ile Leu  
45 1 5 10 15  
Leu Leu Asn Met Leu Ile Ala Leu Met Xaa Arg Asp Arg Gln Gln Cys  
20 25 30  
50 Arg His  
55 (2) INFORMATION FOR SEQ ID NO: 611:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
60

704

- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 611:  
Met Val Phe Gln Gly Phe Ser Ser Ala Phe Cys Leu Ser Ser Thr Ala  
1 5 10 15  
Pro Thr Ser His Pro  
20  
10 (2) INFORMATION FOR SEQ ID NO: 612:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 612:  
Gly Lys Lys Asn Gln Leu Leu Val Ile  
1 5  
20  
25 (2) INFORMATION FOR SEQ ID NO: 613:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 613:  
Met Val Trp Val Leu Trp Ser Ala Pro Ser Leu Ala Pro Pro Trp Val  
1 5 10 15  
30 Gly Pro Cys Trp Pro Ser Thr Gly Asn Cys Cys Leu Cys  
20 25  
35 (2) INFORMATION FOR SEQ ID NO: 614:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 614:  
Met Ala Lys Arg Ser Pro Gly Gly Cys Gly Ser Gly Leu Ile Leu Leu  
45 1 5 10 15  
Cys Cys Gln Pro Cys Arg Pro Thr Ser Ser Ala Pro Met Arg  
20 25 30  
50  
55 (2) INFORMATION FOR SEQ ID NO: 615:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 113 amino acids  
(B) TYPE: amino acid  
60

## (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 615:

5 Ile Thr Ile Ala Ile Gln Met Ile Cys Leu Val Asn Xaa Glu Leu Tyr 15  
1 5 10  
Pro Thr Phe Val Arg Asn Xaa Gly Val Met Val Cys Ser Ser Leu Cys 30  
20 25  
10 Asp Ile Gly Gly Ile Ile Thr Pro Phe Ile Val Phe Arg Leu Arg Glu 45  
35 40  
Val Trp Gln Ala Leu Pro Leu Ile Leu Phe Ala Val Leu Gly Leu Leu 60  
50 55  
15 Ala Ala Gly Val Thr Leu Leu Leu Pro Glu Thr Lys Gly Val Ala Leu 80  
65 70 75  
Pro Glu Thr Met Lys Asp Ala Glu Asn Leu Gly Arg Lys Ala Lys Pro 95  
85 90  
Lys Glu Asn Thr Ile Tyr Leu Lys Val Gln Thr Ser Glu Pro Ser Gly 110  
100 105

25 Thr

30 (2) INFORMATION FOR SEQ ID NO: 616:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 616:

35 Thr Met Lys Asp Ala Glu Asn Leu Gly Arg Lys Ala Lys Pro Lys Glu 15  
1 5 10  
40 Asn Thr

45 (2) INFORMATION FOR SEQ ID NO: 617:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 617:

55 Pro Arg Val Arg Asn Ser Pro Glu Asp Leu Gly Leu Ser Leu Thr Gly 15  
1 5 10  
Asp Ser Cys Lys Leu 20

60

(2) INFORMATION FOR SEQ ID NO: 618:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 618:

5 Gln Ala Asp Asp Leu Gln Ala Thr Val Ala Ala Leu Cys Val Leu Arg 15  
1 5 10  
10 Gly Gly Gly Pro Trp Ala Gly Ser Trp Leu Ser Pro Lys Thr Pro Gly 30  
20 25  
15 Ala Met Gly Gly Asp Leu Val Leu Gly Leu Gly Ala Leu Arg Arg 45  
35 40  
20 Lys Arg Leu Leu 50

(2) INFORMATION FOR SEQ ID NO: 619:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 232 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 619:

25 Glu Gln Glu Lys Ser Leu Ala Gly Trp Ala Leu Val Leu Ala Xaa 15  
1 5 10  
30 Gly Ile Gly Leu Met Val Leu His Ala Glu Met Leu Trp Phe Gly 30  
20 25  
Cys Ser Ala Val Asn Ala Thr Gly His Leu Ser Asp Thr Leu Trp Leu 45  
35 40  
Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr Gly Asp Val Val Pro Gly 60  
50 55  
Thr Met Trp Gly Lys Ile Val Cys Leu Cys Thr Gly Val Met Gly Val 80  
65 70 75  
Cys Cys Thr Ala Leu Leu Val Ala Val Ala Arg Lys Leu Glu Phe 95  
85 90  
50 Asn Lys Ala Glu Lys His Val His Asn Phe Met Met Asp Ile Gln Tyr 110  
100 105  
Thr Lys Glu Met Lys Glu Ser Ala Ala Arg Val Leu Gln Glu Ala Trp 125  
115 120  
55 Met Phe Tyr Lys His Thr Arg Arg Lys Glu Ser His Ala Arg Xaa 140  
130 135  
His Gln Arg Xaa Leu Leu Ala Ala Ile Asn Ala Phe Arg Gln Val Arg 160  
145 150 155

707

- Leu Lys His Arg Lys Leu Arg Glu Gln Val Asn Ser Met Val Asp Ile  
165 170 175
- 5 Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln Asn Leu Ser Ser  
180 185 190
- 10 Ser His Arg Ala Leu Glu Lys Lys Gln Ile Asp Thr Leu Ala Gly Lys Leu  
195 200 205
- Asp Ala Leu Thr Glu Leu Leu Ser Thr Ala Leu Gly Pro Arg Gln Leu  
210 215 220
- 15 Pro Glu Pro Ser Gln Gln Ser Lys  
225 230
- 20 (2) INFORMATION FOR SEQ ID NO: 620:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 620:
- 25 Tyr Gln Ala His His Val Ser Arg Asn Lys Arg Gly Gln Val Val Gly  
1 5 10 15
- 30 Thr Arg Gly Gly Phe Arg Gly Cys Thr Val Trp Leu Thr Gly Leu Ser  
20 25 30
- 35 Gly Ala Gly Lys  
35
- 40 (2) INFORMATION FOR SEQ ID NO: 621:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 57 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 621:
- 45 Leu Gln Cys Glu Ile Cys Gly Phe Thr Cys Arg Gln Lys Ala Ser Leu  
1 5 10 15
- 50 Asn Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser  
20 25 30
- Cys Asn Ile Cys Gly Lys Lys Phe Glu Lys Lys Asp Ser Val Val Ala  
35 40 45
- 55 His Lys Ala Lys Ser His Pro Glu Val  
50 55
- 60 (2) INFORMATION FOR SEQ ID NO: 622:

708

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- 5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 622:
- 10 Ile Thr Ser Thr Asp Ile Leu Gly Thr Asn Pro Glu Ser Leu Thr Gln  
1 5 10 15
- 15 Pro Ser Asp
- 20 (2) INFORMATION FOR SEQ ID NO: 623:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 623:
- 25 Asn Ser Thr Ser Gly Glu Cys Leu Leu Leu Glu Ala Glu Gly Met Ser  
1 5 10 15
- 30 Lys Ser Tyr
- 35 (2) INFORMATION FOR SEQ ID NO: 624:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 51 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 624:
- 40 Cys Ser Gly Thr Glu Arg Val Ser Leu Met Ala Asp Gly Lys Ile Phe  
1 5 10 15
- 45 Val Gly Ser Gly Ser Ser Gly Gly Thr Glu Gly Leu Val Met Asn Ser  
20 25 30
- 50 Asp Ile Leu Gly Ala Thr Thr Glu Val Leu Ile Glu Asp Ser Asp Ser  
35 40 45
- 55 Ala Gly Pro  
50 55
- 60 (2) INFORMATION FOR SEQ ID NO: 625:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 60 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 625:

WO 98/39448

710

Cys Asn Ile Cys Gly Lys Lys Phe Glu Lys Lys Asp Ser Val Val Ala  
 35 40 45  
 5 His Lys Ala Lys Ser His Pro Glu Val Xaa Ile Thr Ser Thr Asp Ile  
 50 55 60  
 Leu Gly Thr Asn Pro Glu Ser Leu Thr Gln Pro Ser Asp Xaa Asn Ser  
 65 70 75 80  
 10 Thr Ser Gly Glu Cys Leu Leu Glu Ala Glu Gly Met Ser Lys Ser  
 85 90 95  
 Tyr Xaa Cys Ser Gly Thr Glu Arg Val Ser Leu Met Ala Asp Gly Lys  
 100 105 110  
 Ile Phe Val Gly Ser Gly Ser Ser Gly Thr Glu Gly Leu Val Met  
 115 120 125  
 20 Asn Ser Asp Ile Leu Gly Ala Thr Thr Glu Val Leu Ile Glu Asp Ser  
 130 135 140  
 Asp Ser Ala Gly Pro Xaa Gln Arg Asp Tyr Ile Cys Glu Tyr Cys Ala  
 145 150 155 160  
 25 Arg Ala Phe Lys Ser Ser His Asn Leu Ala Val His Arg Met Ile His  
 165 170 175  
 Thr Gly Glu Lys His Tyr Xaa  
 180

(2) INFORMATION FOR SEQ ID NO: 629:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 60 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 629:

Gln Tyr Val Arg Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala His  
 1 5 10 15  
 45 Pro Arg Tyr Leu Gln His His Ile Lys Tyr Gln His Leu Lys Lys  
 20 25 30  
 Lys Tyr Val Cys Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu Gln  
 35 40 45  
 Lys Gln Leu Leu Arg His Ala Lys His His Thr Asp  
 50 55 60

(2) INFORMATION FOR SEQ ID NO: 630:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 amino acids
  - (B) TYPE: amino acid

PCT/US98/04493

WO 98/39448

709

Ile Gln Tyr Val Arg Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala  
 1 5 10 15  
 5 His Pro Arg Tyr Leu Gln His His Ile Lys Tyr Gln His Leu Leu Lys  
 20 25 30  
 Lys Lys Tyr Val Cys Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu  
 35 40 45  
 10 Gln Lys Gln Leu Leu Arg His Ala Lys His His Thr  
 50 55 60  
 15 (2) INFORMATION FOR SEQ ID NO: 626:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 626:  
 Asp Gln Arg Asp Tyr Ile Cys Glu Tyr Cys Ala Arg Ala Phe Lys Ser  
 1 5 10 15  
 Ser His Asn Leu Ala Val His Arg Met Ile His Thr Gly Glu Lys  
 20 25 30

(2) INFORMATION FOR SEQ ID NO: 627:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 627:

Arg Ser Ser Arg Ser Lys Thr Gly Ser Leu Gln Leu Ile Cys Lys Ser  
 1 5 10 15  
 Glu Pro Asn Thr Asp Gln Leu Asp Tyr  
 20 25

(2) INFORMATION FOR SEQ ID NO: 628:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 183 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 628:

Leu Gln Cys Glu Ile Cys Gly Phe Thr Cys Arg Gln Lys Ala Ser Leu  
 1 5 10 15  
 Asn Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser  
 20 25 30



711

(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 630:

5 Pro Phe Lys Asp Asp Pro Arg Asp Glu Thr Tyr Lys Pro His Leu Glu  
1 5 10 15  
Arg Glu Thr Pro Lys Pro Arg Arg Lys Ser Gly  
20 25

10 (2) INFORMATION FOR SEQ ID NO: 631:

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 110 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 631:

20 Glu Met Phe Asp Ser Leu Ser Tyr Phe Lys Gly Ser Ser Leu Leu Leu  
1 5 10 15  
Met Leu Lys Thr Tyr Leu Ser Glu Asp Val Phe Gln His Ala Val Val  
20 25 30  
Leu Tyr Leu His Asn His Ser Tyr Ala Ser Ile Gln Ser Asp Asp Leu  
35 40 45  
Trp Asp Ser Phe Asn Glu Val Thr Asn Gln Thr Leu Asp Val Lys Arg  
50 55 60  
Met Met Lys Thr Trp Thr Leu Gln Lys Gly Phe Pro Leu Val Thr Val  
65 70 75 80  
Gln Lys Lys Gly Lys Glu Leu Phe Ile Gln Gln Glu Arg Phe Phe Leu  
85 90 95  
Asn Met Lys Pro Glu Ile Gln Pro Ser Asp Thr Arg Tyr Met  
100 105 110

40 (2) INFORMATION FOR SEQ ID NO: 632:

45 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 632:

50 Leu Glu Lys Val Ala Ser Val Gly Asn Ser Arg Pro Thr Gly Gln Gln  
1 5 10 15  
Leu Glu Ser Leu Gly Leu Leu Ala  
20

60 (2) INFORMATION FOR SEQ ID NO: 633:

712

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 633:

5 Val His Arg Glu Glu Ala Ser Cys Tyr Cys Gln Ala Glu Pro Ser Gly  
1 5 10 15  
Asp Leu  
10

15 (2) INFORMATION FOR SEQ ID NO: 634:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 634:

20 Arg Pro Ala Leu Arg Gln Ala Gly Gly Thr Arg Glu Pro Arg Gln  
1 5 10 15  
Lys Arg Trp Ala Gly Leu  
20

30 (2) INFORMATION FOR SEQ ID NO: 635:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 635:

40 Ala Val Asn Phe Arg Pro Gln Arg Ser Gln Ser Met  
1 5 10

45 (2) INFORMATION FOR SEQ ID NO: 636:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 37 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 636:

50 Met Ile Thr Asp Val Gln Leu Ala Ile Phe Ala Asn Met Leu Gly Val  
1 5 10 15  
Ser Leu Phe Leu Leu Val Val Leu Tyr His Tyr Val Ala Val Asn Asn  
20 25 30  
Pro Lys Lys Gln Glu  
35

60

## (2) INFORMATION FOR SEQ ID NO: 637:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 342 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 637:

10 Glu Glu Met Ala Asp Ser Val Lys Thr Phe Leu Gln Asp Leu Ala Arg  
 1 5 10 15  
 Gly Ile Lys Asp Ser Ile Trp Gly Ile Cys Thr Ile Ser Lys Leu Asp  
 20 25 30  
 Ala Arg Ile Gln Gln Lys Arg Glu Gln Arg Arg Arg Ala Ser  
 35 40 45  
 Ser Val Leu Ala Gln Arg Arg Ala Gln Ser Ile Glu Arg Lys Gln Glu  
 50 55 60  
 Ser Glu Pro Arg Ile Val Ser Arg Ile Phe Gln Cys Cys Ala Trp Asn  
 65 70 75 80  
 Gly Gly Val Phe Trp Phe Ser Leu Leu Phe Tyr Arg Val Phe Ile  
 85 90 95  
 Pro Val Leu Gln Ser Val Thr Ala Arg Ile Ile Gly Asp Pro Ser Leu  
 100 105 110  
 His Gly Asp Val Trp Ser Trp Leu Glu Phe Phe Leu Thr Ser Ile Phe  
 115 120 125  
 Ser Ala Leu Trp Val Leu Pro Leu Phe Val Leu Ser Lys Val Val Asn  
 130 135 140  
 Ala Ile Trp Phe Gln Asp Ile Ala Asp Leu Ala Phe Glu Val Ser Gly  
 145 150 155 160  
 Arg Lys Pro His Pro Phe Pro Ser Val Ser Lys Ile Ile Ala Asp Met  
 165 170 175  
 Leu Phe Asn Leu Leu Leu Gln Ala Leu Phe Leu Ile Gln Gly Met Phe  
 180 185 190  
 Val Ser Leu Phe Pro Ile His Leu Val Gly Gln Leu Val Ser Leu Leu  
 195 200 205  
 His Met Ser Leu Leu Tyr Ser Leu Tyr Cys Phe Glu Tyr Arg Trp Phe  
 210 215 220  
 Asn Lys Gly Ile Glu Met His Gln Arg Leu Ser Asn Ile Glu Arg Asn  
 225 230 235 240  
 Trp Pro Tyr Tyr Phe Gly Phe Gly Leu Pro Leu Ala Phe Leu Thr Ala  
 245 250 255  
 Met Gln Ser Ser Tyr Ile Ile Ser Gly Cys Leu Phe Ser Ile Leu Phe  
 260 265 270

Pro Leu Phe Ile Ile Ser Ala Asn Glu Ala Lys Thr Pro Gly Lys Ala  
 275 280 285  
 Tyr Leu Phe Gln Leu Arg Leu Phe Ser Leu Val Val Phe Leu Ser Asn  
 290 295 300  
 Arg Leu Phe His Lys Thr Val Tyr Leu Gln Ser Ala Leu Ser Ser  
 305 310 315 320  
 Thr Ser Ala Glu Lys Phe Pro Ser Pro His Pro Ser Pro Ala Lys Leu  
 325 330 335  
 Lys Ala Thr Ala Gly His  
 340  
 (2) INFORMATION FOR SEQ ID NO: 638:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 529 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 638:  
 Met Ala Lys Phe Met Thr Pro Val Ile Gln Asp Asn Pro Ser Gly Trp  
 1 5 10 15  
 Gly Pro Cys Ala Val Pro Glu Gln Phe Arg Asp Met Pro Tyr Gln Pro  
 20 25 30  
 Phe Ser Lys Gly Asp Arg Leu Gly Lys Val Ala Asp Trp Thr Gly Ala  
 35 40 45  
 Thr Tyr Gln Asp Lys Arg Tyr Thr Asn Lys Tyr Ser Ser Gln Phe Gly  
 50 55 60  
 Gly Gly Ser Gln Tyr Ala Tyr Phe His Glu Glu Asp Glu Ser Ser Phe  
 65 70 75 80  
 Gln Leu Val Asp Thr Ala Arg Thr Gln Lys Thr Ala Tyr Gln Arg Asn  
 85 90 95  
 Arg Met Arg Phe Ala Gln Arg Asn Leu Arg Arg Asp Lys Asp Arg Arg  
 100 105 110  
 Asn Met Leu Gln Phe Asn Leu Gln Ile Leu Pro Lys Ser Ala Lys Gln  
 115 120 125  
 Lys Glu Arg Glu Arg Ile Arg Leu Gln Lys Lys Phe Gln Lys Gln Phe  
 130 135 140  
 Gly Val Arg Gln Lys Trp Asp Gln Lys Ser Gln Lys Pro Arg Asp Ser  
 145 150 155 160  
 Ser Val Glu Val Arg Ser Asp Trp Glu Val Lys Glu Met Asp Phe  
 165 170 175  
 Pro Gln Leu Met Lys Met Arg Tyr Leu Glu Val Ser Glu Pro Gln Asp

180 185 190  
Ile Glu Cys Cys Gly Ala Leu Glu Tyr Tyr Asp Lys Ala Phe Asp Arg  
195 200 205  
5 Ile Thr Thr Arg Ser Glu Lys Pro Leu Arg Xaa Xaa Lys Arg Ile Phe  
210 215 220  
His Thr Val Thr Thr Thr Asp Asp Pro Val Ile Arg Lys Leu Ala Lys  
225 230 235 240  
10 Thr Glu Gly Asn Val Phe Ala Thr Asp Ala Ile Leu Ala Thr Leu Met  
245 250 255  
15 Ser Cys Thr Arg Ser Val Tyr Ser Trp Asp Ile Val Val Glu Arg Val  
260 265 270  
Gly Ser Lys Leu Phe Phe Asp Lys Arg Asp Asn Ser Asp Phe Asp Leu  
275 280 285  
20 Leu Thr Val Ser Glu Thr Ala Asn Glu Pro Pro Glu Asp Glu Gly Asn  
290 295 300  
Ser Phe Asn Ser Pro Arg Asn Leu Ala Met Glu Ala Thr Tyr Ile Asn  
305 310 315 320  
His Asn, Phe Ser Glu Glu Cys Leu Arg Met Gly Lys Glu Arg Tyr Asn  
325 330 335  
30 Phe Pro Asn Pro Asn Pro Phe Val Glu Asp Asp Met Asp Lys Asn Glu  
340 345 350  
Ile Ala Ser Val Ala Tyr Arg Tyr Arg Ser Gly Lys Leu Gly Asp Asp  
355 360 365  
35 Ile Asp Leu Ile Val Arg Cys Glu His Asp Gly Val Met Thr Gly Ala  
370 375 380  
Asn Gly Glu Val Ser Phe Ile Asn Ile Lys Thr Leu Asn Glu Trp Asp  
385 390 395 400  
Ser Arg His Cys Asn Gly Val Asp Trp Arg Glu Lys Leu Asp Ser Glu  
405 410 415  
45 Arg Gly Ala Val Ile Ala Thr Glu Leu Lys Asn Asn Ser Tyr Lys Leu  
420 425 430  
Ala Arg Trp Thr Cys Cys Ala Leu Leu Ala Gly Ser Glu Tyr Leu Lys  
435 440 445  
50 Leu Gly Tyr Val Ser Arg Tyr His Val Lys Asp Ser Ser Arg His Val  
450 455 460  
Ile Leu Gly Thr Glu Glu Phe Lys Pro Asn Glu Phe Ala Ser Glu Ile  
465 470 475 480  
Asn Leu Ser Val Glu Asn Ala Trp Gly Ile Leu Arg Cys Val Ile Asp  
485 490 495  
60 Ile Cys Met Lys Leu Glu Glu Gly Lys Leu Ile Leu Lys Asp Pro

500 505 510  
Asn Lys Glu Val Ile Arg Val Tyr Ser Leu Pro Asp Gly Thr Phe Ser  
515 520 525  
5 Ser  
505  
10 (2) INFORMATION FOR SEQ ID NO: 639:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 194 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 639:  
15  
20 Lys Lys Arg His Thr Asp Val Glu Glu Phe Tyr Thr Glu Val Gly Glu Ile  
1 5 10 15  
Thr Thr Asp Leu Gly Lys His Glu His Met His Asp Arg Asp Asp Leu  
20 25 30  
25 Tyr Ala Glu Glu Met Glu Arg Glu Met Arg His Lys Leu Lys Thr Ala  
35 40 45  
Phe Lys Asn Phe Ile Glu Lys Val Glu Ala Leu Thr Lys Glu Glu Leu  
50 55 60  
30 Glu Phe Glu Val Pro Phe Arg Asp Leu Gly Phe Asn Gly Ala Pro Tyr  
65 70 75 80  
Arg Ser Thr Cys Leu Leu Glu Pro Thr Ser Ser Ala Leu Val Asn Ala  
85 90 95  
35 Thr Glu Trp Pro Pro Phe Val Val Thr Leu Asp Glu Val Glu Leu Ile  
100 105 110  
40 His Phe Xaa Arg Val Glu Phe His Leu Lys Asn Phe Asp Met Val Ile  
115 120 125  
Val Tyr Lys Asp Tyr Ser Lys Lys Val Thr Met Ile Asn Ala Ile Pro  
130 135 140  
45 Val Ala Ser Leu Asp Pro Ile Lys Glu Trp Leu Asn Ser Cys Asp Leu  
145 150 155 160  
Lys Tyr Thr Glu Gly Val Glu Ser Ser Leu Asn Trp Thr Lys Ile Met Lys  
165 170 175  
50 Thr Ile Val Asp Asp Pro Glu Gly Phe Phe Glu Glu Gly Trp Ser  
180 185 190  
55 Phe Leu  
60 (2) INFORMATION FOR SEQ ID NO: 640:

717

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 640:

5 Arg Ser Gly Leu Gly Leu Gly Ile Thr Ile Ala Phe Leu Ala Thr Leu 15  
 10 Ile Thr Gln Phe Leu Val Tyr Asn Gly Val Tyr Gln Tyr Thr Ser Pro 30  
 15 Asp Phe Leu Tyr Ile Arg Ser Trp Leu Pro Cys Ile Phe Phe Ser Gly 45  
 20 Gly Val Thr Val Gly Asn Ile Gly Arg Gln Leu Ala Met Gly Val Pro 60  
 Glu Lys Pro His Ser Asp 70

## (2) INFORMATION FOR SEQ ID NO: 641:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 101 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 641:

25 Val Thr Gln Pro Lys His Leu Ser Ala Ser Met Gly Gly Ser Val Glu 15  
 30 Ile Pro Phe Ser Phe Tyr Tyr Pro Trp Glu Leu Ala Xaa Xaa Pro Xaa 30  
 35 Val Arg Ile Ser Trp Arg Arg Gly His Phe His Gly Gln Ser Phe Tyr 45  
 50 Ser Thr Arg Pro Pro Ser Ile His Lys Asp Tyr Val Asn Arg Leu Phe 60  
 45 Leu Asn Trp Thr Glu Gly Gln Glu Ser Gly Phe Leu Arg Ile Ser Asn 75  
 Leu Arg Lys Glu Asp Gln Ser Val Tyr Phe Cys Arg Val Glu Leu Asp 90  
 50 Thr Arg Arg Ser Gly 100

## (2) INFORMATION FOR SEQ ID NO: 642:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 233 amino acids  
 (B) TYPE: amino acid

55 Ile Arg His Glu Gln His Pro Asn Phe Ser Leu Glu Met His Ser Lys 15  
 60

718

## (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 642:

5 Met Glu Ala Gln Gln Val Asn Glu Ala Glu Ser Ala Arg Glu Gln Leu 15  
 Gln Xaa Leu His Asp Gln Ile Ala Gly Gln Lys Ala Ser Lys Gln Glu 30  
 10 Leu Glu Thr Glu Leu Glu Arg Leu Lys Gln Glu Phe His Tyr Ile Glu 45  
 Glu Asp Leu Tyr Arg Thr Lys Asn Thr Leu Gln Ser Arg Ile Lys Asp 60  
 15 Arg Asp Glu Glu Ile Gln Lys Leu Arg Asn Gln Leu Thr Asn Lys Thr 80  
 Leu Ser Asn Ser Ser Gln Ser Glu Leu Glu Asn Arg Leu His Gln Leu 95  
 Thr Glu Thr Leu Ile Gln Lys Gln Thr Met Leu Glu Ser Leu Ser Thr 110  
 25 Glu Lys Asn Ser Leu Val Phe Gln Leu Glu Arg Leu Glu Gln Gln Met 125  
 Asn Ser Ala Ser Gly Ser Ser Ser Asn Gly Ser Ser Ile Asn Met Ser 140  
 30 Gly Ile Asp Asn Gly Glu Gly Thr Arg Leu Arg Asn Val Pro Val Leu 160  
 Phe Asn Asp Thr Glu Thr Asn Leu Ala Gly Met Tyr Gly Lys Val Arg 175  
 Lys Ala Ala Ser Ser Ile Asp Gln Phe Ser Ile Arg Leu Gly Ile Phe 190  
 40 Leu Arg Arg Tyr Pro Ile Ala Arg Val Phe Val Ile Ile Tyr Met Ala 205  
 Leu Leu His Leu Trp Val Met Ile Val Leu Leu Thr Tyr Thr Pro Glu 220  
 45 Met His His Asp Gln Pro Tyr Gly Lys 230

## (2) INFORMATION FOR SEQ ID NO: 643:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 643:

55 Ile Arg His Glu Gln His Pro Asn Phe Ser Leu Glu Met His Ser Lys 15  
 60

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

Gly Ser Ser Leu Leu Leu Phe Leu Leu Pro Gln Leu Ile Leu Pro  
20 25 30

Val Cys Ala His Leu His Gln Gln Leu Asn Cys  
35 40

10 (2) INFORMATION FOR SEQ ID NO: 644:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 63 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 644:

15 Ser Phe Phe Ile Ser Gln Gln Lys Gly His Leu Leu Gln Ala Gln  
1 5 10 15

20 Arg His Pro Trp Val Ala Gly Ala Leu Val Gly Val Ser Gly Gly Leu  
20 25 30

25 Thr Leu Thr Thr Cys Ser Gly Pro Thr Gln Lys Pro Ala Thr Lys Asn  
35 40 45

Tyr Phe Leu Lys Arg Leu Leu Gln Gln Met His Ile Arg Ala Asn  
50 55 60

30

35

A. The indications made below relate to the microorganism referred to in the description  
on page 116, line N/A

## B. IDENTIFICATION OF DEPOSIT

Further deposits are identified on an additional sheet ☐

Name of depositary institution

American Type Culture Collection

Address of depositary institution (including postal code and country)

12301 Parklawn Drive  
Rockville, Maryland 20852  
United States of America

Date of deposit February 26, 1997

Accession Number 97897

## C. ADDITIONAL INDICATIONS (leave blank if not applicable)

This information is continued on an additional sheet ☐

## D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)

## E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")

For receiving Office use only

For International Bureau use only

☒ This sheet was received with the International application☐ This sheet was received by the International Bureau on:

Authorized officer

Susan White

PCT International Division

Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 116, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209043	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

<input checked="" type="checkbox"/> This sheet was received with the international application	For receiving Office use only
Authorized officer Susan White PCT International Division	For International Bureau use only
Authorized officer	This sheet was received by the International Bureau on: <input type="checkbox"/>

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 119, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit September 4, 1997 Accession Number 209235	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

<input checked="" type="checkbox"/> This sheet was received with the international application	For receiving Office use only
Authorized officer Susan White PCT International Division	For International Bureau use only
Authorized officer	This sheet was received by the International Bureau on: <input type="checkbox"/>

WO 98/39448

723

PCT/US98/04493

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 122, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97898
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

WO 98/39448

724

PCT/US98/04493

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 122, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209044
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13biv)

A. The indications made below relate to the microorganism referred to in the description on page 126, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit February 26, 1997 Accession Number 97899	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer Susan White PCT International Division		Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13biv)

A. The indications made below relate to the microorganism referred to in the description on page 126, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209045	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer Susan White PCT International Division		Authorized officer	



## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>130</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit April 28, 1997	Accession Number 209011
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>131</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97900
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b<sub>2</sub>)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 26, 1997
Accession Number	97901
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer		Authorized officer	
Susan White PCT International Division			

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b<sub>1</sub>)

A. The indications made below relate to the microorganism referred to in the description on page 131, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	May 15, 1997
Accession Number	209046
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer		Authorized officer	
Susan White PCT International Division			

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b1b)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country)	
12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	May 15, 1997
Accession Number	209047
C. ADDITIONAL INDICATIONS (leave blank if not applicable)	
This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b1b)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country)	
12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	May 22, 1997
Accession Number	209076
C. ADDITIONAL INDICATIONS (leave blank if not applicable)	
This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b(ii))

A. The indications made below relate to the microorganism referred to in the description on page 140, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit August 21, 1997	Accession Number 209215
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer Susan White PCT International Division		Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b(ii))

A. The indications made below relate to the microorganism referred to in the description on page 160, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97904
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer Susan White PCT International Division		Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page <u>154</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT	Further deposits are identified on an additional sheet <input type="checkbox"/>
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit July 3, 1997	Accession Number 209139
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
<div style="display: flex; justify-content: space-between;"> <div> <p>For receiving Office use only</p> <p><input checked="" type="checkbox"/> This sheet was received with the international application</p> <p>Authorized officer Susan White PCT International Division</p> </div> <div> <p>For International Bureau use only</p> <p><input type="checkbox"/> This sheet was received by the International Bureau on:</p> <p>Authorized officer</p> </div> </div>	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page <u>153</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT	Further deposits are identified on an additional sheet <input type="checkbox"/>
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209049
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
<div style="display: flex; justify-content: space-between;"> <div> <p>For receiving Office use only</p> <p><input checked="" type="checkbox"/> This sheet was received with the international application</p> <p>Authorized officer Susan White PCT International Division</p> </div> <div> <p>For International Bureau use only</p> <p><input type="checkbox"/> This sheet was received by the International Bureau on:</p> <p>Authorized officer</p> </div> </div>	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 153, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit February 26, 1997 Accession Number 97903 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer Susan White PCT International Division		Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 142, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit June 12, 1997 Accession Number 209119 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer Susan White PCT International Division		Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 146, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97902
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 146, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209048
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 160, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209050 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 142, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit February 12, 1998 Accession Number 209627 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer



*What Is Claimed Is:*

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
  - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
    - (f) a polynucleotide which is a variant of SEQ ID NO:X;
    - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
    - (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
    - (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
  2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.
  3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.
8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.
9. A recombinant host cell produced by the method of claim 8.
10. The recombinant host cell of claim 9 comprising vector sequences.
11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:
  - (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;
  - (c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

745

746

(g) a variant of SEQ ID NO:Y;

(h) an allelic variant of SEQ ID NO:Y; or

(i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length comprises sequential amino acid deletions from either the C-terminus or the

body that binds specifically to the isolated polypeptide of

expresses the isolated polypeptide of claim

comprising:

under conditions such that

said polypeptide is expressed; and

(b) recovering said polypeptide.

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and  
(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO:X in a cell;  
(b) isolating the supernatant;  
(c) detecting an activity in a biological assay; and  
(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 22.